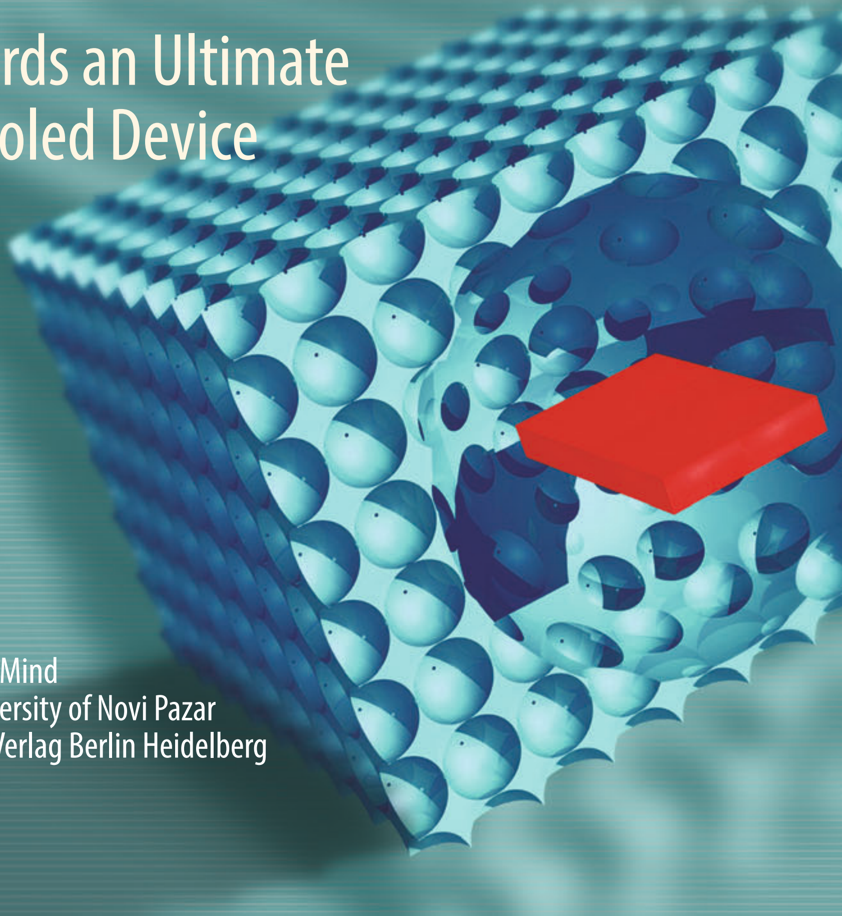
Zoran Jakšić

Micro and Nanophotonics for Semiconductor Infrared Detectors

Towards an Ultimate Uncooled Device



Academic Mind
State University of Novi Pazar
Springer-Verlag Berlin Heidelberg
2014

Micro and Nanophotonics for Semiconductor Infrared Detectors

Zoran Jakšić

Micro and Nanophotonics for Semiconductor Infrared Detectors

Towards an Ultimate Uncooled Device

AKADEMSKA MISAO
ACADEMIC MIND

Springer

Zoran Jakšić
Centre of Microelectronic Technologies,
Institute of Chemistry, Technology and
Metallurgy
University of Belgrade
Belgrade
Serbia

ISBN 978-3-319-09673-5 ISBN 978-3-319-09674-2 (eBook)
ISBN 978-86-7466-490-2
DOI 10.1007/978-3-319-09674-2

Library of Congress Control Number: 2014949336

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The advances in micro and nanotechnologies nowadays are nothing less than spectacular. Materials, structures, devices, and systems never seen before are appearing at an accelerated pace, while new properties and novel phenomena are being discovered almost every day. The microelectromechanical system (MEMS) and microoptoelectromechanical (MOEMS) technologies have brought a multitude of devices with ultracompact dimensions and with vastly enhanced or even completely novel properties and characteristics. Nanotechnologies are continuously expanding and improving the toolbox of available materials and systems, smudging the limits between technological and natural, even biological, and the potentials for further expansion appear literally limitless. The quest toward ever smaller dimensions and ever-increasing control over structures and properties is continuing and is already having a profound influence on practically all fields of applied science, steadily permeating more and more facets of ordinary life.

What is valid generally, is also valid in optoelectronics and photonics, actually maybe even more so than in some other fields. Completely novel areas are being created, ensuring manipulation with light at an unprecedented level. Micro and nanofabrication of structures with a periodicity of refractive index that may extend to all three dimensions have resulted in photonic bandgap materials, also known as photonic crystals, ensuring manipulation with photons in a manner similar to that with electrons in semiconductors. Subwavelength structuring of metal-dielectric materials brought metamaterials, structures possessing electromagnetic properties not readily met in nature and often unexpected and sometimes even counterintuitive. These include materials with values of refractive index below zero or equal to zero. It also created the field of plasmonics, dedicated to manipulation with evanescent, surface-bound electromagnetic waves that build a link between electronics and optics, ensuring devices with compact dimensions of micro and nanoelectronic circuits that operate at the speed of optical signals. Structures and devices are being created which until recently were deemed impossible.

The advent of micro and nanophotonics has brought astounding enhancements to different passive and active devices. In the field of infrared technologies, MEMS and NEMS, it ensured the appearance of a new generation of thermal detectors, with properties vastly surpassing traditional devices. Novel nanodevices are being produced where quantum structures are used to reach new levels of performance.

However, there is a glaring gap in the list of infrared devices enhanced by the micro and nanophotonic technologies. For decades, simple intrinsic photonic (semiconductor) detectors have been far superior to other devices for infrared detection. Even today they remain the optimum choice if one simultaneously requires sensitivity, selectivity, and speed at an acceptable cost. Yet relatively little has been done to improve their performance using the same technologies that brought the renaissance to so many other fields. There is no obvious reason why micro and nanophotonics should not bring about similar improvements to photonic devices that were superior to begin with. The results actually do exist, but are scattered throughout the literature and are sometimes quite underwhelming. The author is not aware of a systematic treatise dedicated to the application of novel approaches stemming from microphotonic and nanophotonic technologies to traditional semiconductor infrared detectors. The goal of this book is to fill this gap.

Nowadays, two large classes of optoelectronic devices are used for detection of mid- (MWIR) and long-wavelength (LWIR) infrared radiation in the atmospheric windows (3–5) and (8–14) μm: photonic (quantum) detector and thermal detectors. Photonic detectors have been ruling the market for several decades, mostly of intrinsic type, based on indium antimonide and mercury cadmium telluride. Besides having high values of specific detectivity and generally high performance, they are characterized by high response speeds. Even today they largely outperform the novel generation of thermal infrared detectors.

The largest drawback of photonic detectors of infrared radiation compared to thermal devices is their requirement for cooling. For this purpose one typically uses Dewar flasks with liquid nitrogen or various kinds of active coolers, including Joule–Thompson cryostats or multistage thermoelectric coolers. Such cooling systems obviously make the detector units overly complex, impractical for handling, and vastly increase their price.

It is no wonder that through the decades a large part of the research and development connected with semiconductor detector elements for mid- and long-wavelength infrared radiation was dedicated exactly to the question of how to fabricate devices with high specific detectivities and high speeds at elevated temperatures (the “HOT”—Higher Operating Temperature detectors). The seemingly simple task of fabricating uncooled infrared detectors with background-limited operation engaged armies of researchers and for some of them became the holy grail of photonic detector design.

This book is dedicated to the same task, observed however through the prism of modern microphotonic and nanophotonic technologies. We offer a systematic approach to different available methods for increase in the operating temperatures of photonic detectors up to room temperature, while simultaneously retaining the performance of cryogenically cooled devices. The avoidance of cooling requirements would reduce the detector price to a fraction of the previous one, but the main gain should be expended in improved performance, higher robustness and ease of use, and much more widespread utilization.

It is necessary to mention that this text is dedicated solely to intrinsic photonic infrared detectors. This means that we do not consider more complex structures of detector materials themselves, which include different nanocomposites like quantum wells, quantum wires, and quantum dots. The topic of this work is micro and nanophotonic technologies, and not advanced detector materials, although they also represent a huge field for potential improvements. However, practically all of the methods and techniques described here are also applicable to devices based on advanced detector materials.

An important motivation for this kind of book is obviously the wide and ever expanding applicability of infrared detectors, especially those for the long-wavelength range. Naturally, in today's society all kinds of sensors are in increasing demand, but infrared devices assume a special position.

The situation was quite different in the past. The majority of the MWIR and LWIR detectors were used for military applications, for implementations like night vision, heat seeking and guidance, etc. Materials used for these detectors were also specific and were mostly limited to narrow-bandgap semiconductors. These detectors were also used for lab work and generally scientific applications in physics, meteorology, geology, material science, astronomy, remote sensing, airspace research, etc. Mostly their users were highly demanding. This explains why cooled detectors became widespread—for such users the price was most often, albeit not always, second to performance.

Usually, a large field of use for MWIR and LWIR detectors is overlooked—the free-space optical (FSO) telecommunications. This is certainly unusual when one bears in mind that communications themselves are one of the pillars of civilization itself, actually one could maintain that communications were those that created civilization in the first place.

In the field of free-space communication a special position belongs to microwave and optical wavelength ranges. In highly urbanized environments the so-called last-mile solutions are becoming increasingly important—instead of having to build transmission lines into the already existing infrastructure, which often requires astronomical expenses, it is much easier to utilize various wireless approaches. Since the majority of the existing FSO systems belong to the near infrared range, the available bands are becoming quite cramped. The long-wavelength IR range is thus mentioned as an alternative. This range is less sensitive to atmospheric disturbances like fog, smoke, rain, etc.; LWIR wavelengths are much less likely to accidentally damage eyes since vastly higher radiation thresholds are tolerated in this range. Last but not least, the LWIR range is not yet covered by strict regulations: it is literally a new territory waiting to be conquered. Low-cost uncooled detectors with high performance would be a significant step toward a more widespread use of such systems.

The applicability of the most of the methods presented in this book does not end with infrared detectors and is in reality much wider. An obvious and important field of use of different micro and nanophotonic methods for detection improvement elaborated in this book are solar cells, especially thin-film photovoltaics. Solar energy harvesting represents a wide area of its own, but many of the approaches presented here may be applied to it with little or no modifications.

This book is organized into three chapters. The first chapter is introductory and considers various demands that may be posed to a general infrared detector, as well as which real types of detectors satisfy these demands best. General figures of merit of infrared detectors are analyzed and stress is given to the product of specific detectivity and bandwidth, which is taken to be one of the key factors of detector quality. The work then proceeds to present various particular mechanisms that define the mentioned figures of merit, especially the prevailing generation-recombination processes and the related noise mechanisms. The goal of this chapter is to estimate the most convenient choice among different infrared detectors and to consider which of its figures of merit could be optimized and how.

The second chapter of the book analyzes the possibility to enhance the detector performance utilizing photon management or light management. These methods may be denoted as equilibrium ones, since they do not generate excess charge carrier concentrations for their function and are related solely to handling of light flux outside the detector and within it. The equilibrium methods include various light trapping schemes both in the far field and in the near field with a goal to enhance the detector response. Since there are numerous approaches to this kind of enhancement, after a short literature review a possible classification of these approaches is proposed. Besides including all photon management methods known to the author, the use of the classification points out to some possible novel approaches, as well as to the use of certain methods that until now were utilized only in other fields of optical engineering.

Photon management methods are divided into two large groups. One group comprises structures that increase the incident optical flux. Among these, the text first considers optical concentrators, which include refractive and diffractive lenses, either separate or integrated with the detector, as well as some reflective nonimaging concentrators. The next subgroup is structures for decrease in reflection at the incident plane, which include interference antireflection films and diffractive subwavelength antireflection structures. An interesting new field in this domain is the use of plasmonics which brings optical concentration to the micro and nano levels—the enhancement occurs in the near field and utilizes evanescent waves.

The second equilibrium group encompasses structures for increase in the optical path of the beam which already entered the active area of the detector, the so-called light trapping structures. These structures simultaneously increase radiative lifetime through the mechanism of reabsorption—photon recycling. They include different surface relief structures for the increase in total internal reflection, from diffractive to macroscopic ones. Reflective detector surfaces also belong to this group, both the back-side ones and full resonant cavities (RCE—resonant cavity enhancement) with reflective surfaces both of the front and the back side of the detector. The most advanced structures for optical path and radiative time increase are radiative shields and photonic crystal enhancement structures, which represent a full cavity enhancement and may support the existence of multiple modes.

The third chapter of the book considers nonequilibrium methods for thermal noise management, which include manipulation with charge carrier concentration within the detectors utilizing external fields and which are implemented through the application of general micro and nanosystem techniques. These are very important, since they attack the very reason why photonic detectors are cooled. Since the operating wavelengths are long, i.e., the energies of incident photons are low, the bandgap of the utilized material must also be small. This means that thermal processes of carrier generation will start to compete with the detection process itself and thus cause excessive increase of generation-recombination noise. This is especially valid for noise caused by Auger processes that prevail in sufficiently pure semiconductors. The goal of nonequilibrium methods is to cause controlled perturbation of carrier distribution within the detector and thus locally decrease the carrier concentration below its equilibrium value, thereby directly decreasing the level of total generation-recombination noise. A systematic approach to these methods and a newly proposed classification stemming from it again show that in addition to the methods published in the literature one could utilize some new ones.

Special care is dedicated to the consideration of properties common to all types of nonequilibrium devices. Based on this a general model is defined to be utilized in simulation of any nonequilibrium detector, including the novel ones. The model is derived for isothermal processes, but can be extended to nonisothermal ones. When deriving the model, special care is dedicated to the limits of the model and the generality of the introduced approximations.

The third chapter considers the use of contact phenomena, i.e., structures utilizing a combination of a built-in dopant concentration gradient and the external electric field for suppression of Auger processes. Two main types are considered, exclusion-based photoconductors and extraction-based photodiodes. The text further considers the use of crossed electric and magnetic field and analyzes the use of galvanomagnetic phenomena, i.e., the magnetoconcentration effect for suppression of Auger processes. Finally, conclusions are drawn and assessment of possible future directions is given.

Contents

1 Introduction: A Path to an Ideal Photonic Infrared Detector 1
  1.1 Infrared Detector Performance and Main Figures of Merit 1
    1.1.1 Some Definitions 1
    1.1.2 Photodetector Figures of Merit 2
    1.1.3 "Ideal" IR Detector 2
  1.2 Classification of IR Detectors 6
    1.2.1 Thermal Detectors 6
    1.2.2 Photoemissive Vacuum Detectors 8
    1.2.3 Detectors with Interaction of Electromagnetic Radiation 8
    1.2.4 Photonic Detectors 9
    1.2.5 Choice of Optimum Photodetector 11
  1.3 On Specific Detectivity of Photonic Infrared Detectors 12
  1.4 Recombination Mechanisms in Direct Narrow-Bandgap Semiconductor 16
    1.4.1 Radiative Processes 17
    1.4.2 Auger Processes 22
    1.4.3 Shockley-Read Processes 29
    1.4.4 Optical Generation 32
    1.4.5 Influence of Background Radiation 33
    1.4.6 Total G-R Rate of Narrow-Bandgap Semiconductors 33
  1.5 Noise in Semiconductor IR Detectors 34
    1.5.1 Thermal Noise 35
    1.5.2 Generation-Recombination Noise 35
    1.5.3 1/f Noise 36
    1.5.4 Total Detector Noise 36
  1.6 Optimizing D^*F^* Product 38
    1.6.1 Conditions to Reach Maximum D^*f^* in Photonic Detectors 38
    1.6.2 Generalized Photonic IR Detector 40

2 Photon Management 43
2.1 Fundamentals of Photon Management in Photodetectors 43
2.1.1 Introduction 43
2.1.2 Fundamental Limits of Photon Management 45
2.2 Nonimaging and Imaging Optical Concentrators 47
2.3 Refractive Concentrators 49
2.3.1 Immersion Microlenses 49
2.3.2 MEMS Fabrication of Infrared Microlenses 51
2.4 Reflective Concentrators 54
2.4.1 CPC (Winston Collector) 55
2.4.2 Parabolic Reflector 57
2.4.3 Conic Concentrator 59
2.4.4 Hyperbolic (Trumpet) Geometry 59
2.4.5 Other Reflector Geometries 60
2.4.6 Fresnel Reflector 60
2.5 Diffractive Optical Concentrators 61
2.5.1 Definitions 61
2.5.2 Fresnel Lenses 64
2.5.3 Holographic Optical Elements 68
2.6 Antireflection Structures 70
2.6.1 Antireflection Interference Films 71
2.6.2 Gradient-Index Single-Layer Dielectric Films 76
2.6.3 Antireflective Diffractive Structures 77
2.6.4 Bionic Approach: Moth-Eye Structures 83
2.6.5 Random Surface Corrugations 84
2.6.6 Subsurface Scatterers 85
2.7 Metamaterials as Antireflection Structures 86
2.7.1 Metamaterial-Containing Antireflection Multilayers 88
2.7.2 Metamaterial Superabsorbers 89
2.8 Light Trapping by Optical Path Increase 90
2.9 High-Reflection Structures 92
2.9.1 Introduction 92
2.9.2 Specular Reflectivity 93
2.9.3 Metal Mirrors 94
2.9.4 Dielectric Interference Mirrors 94
2.9.5 Omnidirectional Reflectors 98
2.9.6 Metal-Dielectric Mirrors 100
2.9.7 Retroreflectors 101
2.9.8 Diffuse Mirrors 101
2.10 TIR Structures 103
2.11 Resonant Cavity Enhancement (RCE) 105
2.12 Radiative Shields 110
2.13 Photonic Crystal Enhancement (PCE) 112
2.14 Radiative Lifetime Increase 118

2.15 Plasmonics for Light Trapping 120
2.16 Optical Antennas 123
2.17 Plasmonic Nanoparticles and Redshifting 125

3 Charge Carrier Management (Thermal Noise Engineering) 129
3.1 Nonequilibrium Suppression of Auger Generation-Recombination 129
3.1.1 Physical Principles 129
3.1.2 Some General Properties of Nonequilibrium Photodetectors 132
3.2 Limits of Nonequilibrium Detector Operation 133
3.2.1 Lower Limit of Operation of Nonequilibrium Devices 133
3.2.2 Maximum Fields for Nonequilibrium Suppression 135
3.3 General Model of Nonequilibrium Photodetectors with Auger Suppression 140
3.3.1 Poisson Equation 140
3.3.2 Continuity Equation 141
3.3.3 Transport Equation (Current Density in Nonequilibrium Detectors) 142
3.3.4 One-Dimensional Model of Nonequilibrium Photodetector 149
3.4 Contact Phenomena and Galvanic Suppression of Auger Processes 151
3.4.1 Introduction and Physical Concept 151
3.4.2 Mathematical-Physical Model of Nonequilibrium Detectors Based on Minority Carrier Exclusion and/or Extraction: Boundary Conditions 154
3.5 Exclusion Photoconductors 156
3.5.1 History of the Utilization of Minority Carrier Exclusion 157
3.5.2 Various Implementations of Exclusion Photoconductors 158
3.5.3 Approximate Analytic Calculation of Exclusion Detector Parameters 159
3.5.4 Carrier Distribution in Exclusion Detector 163
3.5.5 g-r Rates in Exclusion Photodetector 166
3.5.6 Current-Voltage Characteristics of Exclusion Photoconductor 168
3.5.7 Calculated Noise of Exclusion Photoconductors 170
3.5.8 Sensitivity of Exclusion Photoconductor 170
3.5.9 Detectivity of Exclusion Photoconductor 171
3.5.10 Response Time of an Exclusion Photoconductor 172

3.5.11 Experimental Results on Excluded Nonequilibrium Photodetectors. 173
3.5.12 Advantages and Shortcomings of Exclusion-Based Nonequilibrium Photodetectors 174
3.6 Extraction Photodiodes 175
3.6.1 Historical Development of the Use of Minority Carrier Extraction in Semiconductor Detectors 177
3.6.2 Various Implementations of Nonequilibrium Photodetectors Utilizing Extraction and Exclusion 178
3.6.3 Approximate Analytical Calculation of Narrow-Bandgap Extraction-Exclusion Photodiodes with Abrupt Junction 181
3.6.4 Carrier Distribution in Extraction-Exclusion Photodiode 184
3.6.5 g-r Rates in Extraction-Exclusion Photodiodes 186
3.6.6 Current-Voltage Characteristics of Extraction-Exclusion Photodetector with Homojunction 188
3.6.7 Photoelectric Gain and Noise of Extraction Photodiode 190
3.6.8 Sensitivity and Detectivity of Extraction-Exclusion Detectors 193
3.6.9 Experimental Results on Extraction-Exclusion Detectors 194
3.6.10 Advantages and Disadvantages of Extraction-Exclusion Devices 196
3.7 Galvanomagnetic Methods—Magnetoconcentration Photodetector 196
3.7.1 Physical Concept. 196
3.7.2 History of Magnetoconcentration Effect Research 198
3.7.3 Model of Nonequilibrium Magnetoconcentration Detector 199
3.7.4 Analytical Consideration of Nonequilibrium Magnetoconcentration Photodetector 201
3.7.5 Carrier Distribution in Nonequilibrium Magnetoconcentration Detector 205
3.7.6 Recombination Rates in Nonequilibrium Magnetoconcentration Photodetector 207
3.7.7 Current-Voltage Characteristics of Nonequilibrium Magnetoconcentration Photodetectors. 209
3.7.8 Noise in Nonequilibrium Magnetoconcentration Photodetectors. 212
3.7.9 Sensitivity and Detectivity of Nonequilibrium Magnetoconcentration Detectors 214

3.7.10 Spectral Characteristics of Nonequilibrium Magnetoconcentration Detectors . . . 216
3.7.11 Response Time of Nonequilibrium Magnetoconcentration Photodetector . . . 217
3.7.12 Experimental Results on Nonequilibrium Magnetoconcentration Photodetectors . . . 219
3.7.13 Advantages and Shortcomings of Nonequilibrium Magnetoconcentration Photodetectors . . . 222
3.7.14 Basic Implementations of Nonequilibrium Magnetoconcentration Photodetectors . . . 223
3.8 Hybrid Methods . . . 224
3.8.1 Magnetoconcentration-Exclusion Device . . . 224
3.8.2 Magnetoconcentration-Extraction Device . . . 225
3.8.3 Dember Magnetoconcentration Detector . . . 226
3.9 On Application of Nonequilibrium Detectors: LWIR Free-Space Optics . . . 228

Conclusion . . . 231

References . . . 235

Index . . . 253

Notation

a	Basis of the unit cell of a PBG material
A	1. Photodetector active area 2. RCE detector enhancement factor due to interference
A_0	Optical active area of photodetector
A_e	Electric active area of photodetector
B	Magnetic induction
c	Speed of light in vacuum
C_m	Integer index, $m = 1, 2, 3, \ldots$
C_{SR}	Mean probability of electron capture by a trap
D	1. Electric induction 2. Detectivity 3. Diffusion coefficient
d	Thickness of detector active area
D^*	Specific detectivity
D^{**}	Specific detectivity per spatial angle
D^*_λ	Specific detectivity at a given wavelength
D_a	Ambipolar diffusion coefficient
D_n, D_p	Diffusion coefficient of electrons (n), holes (p)
E	Photon energy
E_0	Electric field of a plane wave
E_c	Energy at the bottom of the conduction band
E_f	Fermi level in equilibrium
E_{fn}, E_{fp}	Quasi-Fermi level of electrons, holes
E_g	Semiconductor bandgap
E_{geff}	Effective bandgap
E_{krit}	Critical electric field in exclusion
E_{MB}	Burstein-Moss shift due to band filling
E_T	Threshold energy
E_t	Trap energy level

E_v	Energy corresponding to the top of the valence band
$E_{x,y}$	Electric field in x, y direction
F	Distribution function
F	1. Force caused by an external field 2. Mathematical function
$F_{1,2\ldots}$	Bloch function overlap integral
f^*	Frequency range of specific detectivity
f_0	Fermi-Dirac distribution function in equilibrium
f_1	Nonequilibrium part of the distribution function
$f_{3\mathrm{db}}$	Cutoff frequency
f_F	Focal length
f_s	Optical grating filling factor
f_t	Trap occupation probability
g_A	Total (net) rate of Auger generation-recombination
G_A	Auger generation rate coefficient
G_A^i	Auger generation rate in intrinsic semiconductor
g_{BG}	Generation rate due to background radiation
G_{opt}	Optical generation rate
G_r	Radiative generation rate
G_{SR}	Shockley-Read generation rate
g_{tot}	Total generation-recombination rate
h	Planck constant, 6.625×10^{-34}Js
$\hbar$	$h/2\pi$
H	Magnetic field
H_d	Total width of photodiode
H'	Width of the photodiode p-region
I	Photodetector current
$I(x,y)$	Spatial distribution of optical radiation intensity
I_d	Dark current
I_f	Photosignal current
I_n	Noise current
$i_{g\text{-}r}$	g-r noise current
i_J	Johnson–Nyquist noise current
$i_{1/f}$	1/f noise current
i_r	Radiative generation rate
j	Integer number
J	Current density
J_S	Saturation current density
K	1. Wave vector 2. Integer number
k_b	Boltzmann constant
K	Numerical constant
	Stratification vector

K_i	Kinetic coefficient of carrier transport connected with mechanism i
K_m	Modified Bessel function of second kind, m-th order
l	1. Length of detector active area 2. Integer number
L_a	Ambipolar diffusion length
L_{ekskl}	Exclusion zone length
L_n, L_p	Diffusion length of electrons (n), holes (p)
m	Integer number
M	Integer number
m_{0_*}	Free electron rest mass
m_{n_*}, m_p	Effective mass of electrons (n), holes (p)
n	Refractive index
n_e	Electron concentration
N_D^+	Ionized donor concentration
n_0	Electron equilibrium concentration
n_1	Electron equilibrium concentration for $E_t = E_f$
N_A	Acceptor concentration
N_A^-	Ionized acceptor concentration
N_D	Donor concentration
NEP	Noise equivalent power
n_i	Intrinsic concentration
N_t	Trap concentration
N_Δ	Absolute value of difference of acceptor and donor concentration $\lvert N_D - N_A \rvert$
p	Hole concentration
p_0	Hole intrinsic concentration
P_0	Incident radiation power
p_1	Hole intrinsic concentration for $E_t = E_f$
P_1	Kane matrix element
P_b	Dissipation power
P_e	Photon emission probability
P_{ij}	Transfer matrix element
p_{nb}, p_{ni}	Coefficients of transport equation
q	Electron elementary charge
R	1. Optical reflection coefficient 2. Electric resistance
r	Position vector
r'_A	Auger recombination rate
R_d	Dynamic resistance
r_d	Radius of a circular active area of detector
R_F	Radius of the outermost ring of a Fresnel lens
R_I	Current responsivity of a photodetector
R_m	Fresnel zone diameter

r_m	Fresnel reflectance of the structure m
r_{mc}	Empirical coefficient of the diffusion part in magnetoconcentration detector
R_{SR}	Shockley-Read recombination rate
R_v	Voltage responsivity of photodetector
r_w	White's empirical coefficient of the diffusion current part
R_λ	Responsivity at a given wavelength
S	Surface recombination rate
SW	Standing wave ratio
T	Absolute temperature in K
t	Time
T_D	Debye temperature
T_d	Detector temperature
T_{if}	Matrix element of transit between an initial (i) and final (f) state
T_{tr}	Transit time
u	Coefficient
U	Bias voltage
U_j	Chebyshev polynomial of the second kind, j-th order
V	Volume
$V(r)$	Position-dependent potential
V_n	Noise voltage
v_{tn}, v_{tp}	Thermal velocity of electrons (n), holes (p)
w	1. Photodetector active area width 2. Depletion area width in photodiode
W	Photon absorption probability per unit time charge carrier scattering probability per unit time
x_{Cd}	Cadmium molar fraction in $\mathrm{Hg_{1-x}Cd_xTe}$
x	x-coordinate
X	Vector term of nonequilibrium distribution function
x_j	x-coordinate of p-n junction
y	y-coordinate
z	z-coordinate
α	Absorption coefficient
$\alpha_{1/f}$	$1/f$ (flicker) noise constant
α_n	Capture coefficient of electron at trap
β	Coefficient
δ	Kronecker's delta
δf_t	Perturbation function of trap filling
Δ	Small spatial step
ΔE_i	Bandwidth shift caused by the interaction i
Δf	Frequency range
$\Delta n, \Delta p$	Excess concentration of electrons (n), holes (p)
ε	Dielectric permittivity
ε_0	Dielectric permittivity of vacuum

ε_s	Static dielectric permittivity
ε_∞	High-frequency dielectric permittivity
$\varepsilon', \varepsilon_{1,2}$	Effective emissivity
Φ	Photon flux density
Φ_b	Background photon flux density
γ	1. Ratio of intrinsic Auger lifetimes for processes 1 and 7 2. Position of defect mode in a PBG structure
Γ	Photoelectric gain
η	1. Quantum efficiency of photodetector 2. Diffractive efficiency of holographic optical element
η_F	Fresnel zone plate efficiency
k	Bloch wavenumber
λ	Wavelength
λ_{co}	Cutoff wavelength
λ_D	Debye length of screened potential
Λ	Step of a diffractive grating
$\mu_{\mathrm{eb,hb,t}}$	Coefficient
μ_n, μ_p	Electron, hole mobility
μ_H	Hall mobility of charge carriers
ν	Oscillator frequency
ν_{co}	Cutoff frequency
θ	Angle
ρ	1. Volume density of photon energy 2. Charge carrier density
σ	Electric conductivity
σ_n, σ_p	Effective cross-section of electron (n) or hole (p) scattering on a trap
τ	Charge carrier lifetime
τ_A	Auger lifetime
τ_A^i	Intrinsic Auger lifetime
τ_r	Radiative lifetime
τ_r^i	Intrinsic radiative lifetime
τ_{SR}	Shockley-Read lifetime
τ_{n0}, τ_{p0}	Shockley-Read lifetime for completely empty ($n0$) or completely filled ($p0$) traps
τ_R	Response time
ω	Angular frequency
ω_0	Angular frequency corresponding to the knee of $1/f$ noise dependence
ξ	Coefficient
ψ	Phase part of Fresnel reflectance

Chapter 1
Introduction: A Path to an Ideal Photonic Infrared Detector

1.1 Infrared Detector Performance and Main Figures of Merit

1.1.1 Some Definitions

An infrared (IR) detector is a device converting the energy of infrared radiation into electrical signals, visible light, or some other measurable physical quantity. Its purpose is to enable the use of human senses to observe the presence of infrared radiation and quantitatively determine its intensity. The detector may detect the picture as a whole ("imaging" detector, e.g., night vision systems) or only a part of the picture with sufficiently small dimensions to have a homogeneous radiation distribution over its surface ("point" detector, or detector element—basically, a "non-imaging" detector). Since any picture detector may in principle be reduced to a matrix of point detectors, in this text we consider only detector elements and denote them simply as detectors.

Infrared radiation is electromagnetic radiation with wavelengths ranging from 0.75 μm (the upper limit of the red part of the visible spectrum) to 1,000 μm (lower wavelength limit of microwave radiation range). According to an existing classification [1] the infrared range is divided into several subranges, and the main ones are near-infrared radiation (NIR or NWIR, the range (0.75–3) μm), middle (MIR or MWIR, (3–6) μm), far or long-wavelength infrared radiation (LWIR or FIR, (6–15) μm), and extreme infrared (15–1,000) μm). In each of the mentioned ranges there is at least one transmission maximum in the spectral transmittance of the Earth's atmosphere (atmospheric window).

This text handles the detector elements for the detection of infrared radiation in the atmospheric windows (3–5) μm (MWIR) and (8–14) μm (LWIR). Only intrinsic semiconductor devices are considered.

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2_1

1.1.2 Photodetector Figures of Merit

In an ideal case the operation of a photodetector is fundamentally limited by the discrete (photonic) nature of the optical radiation, i.e., by the fluctuations of the number of photons in the detected incident signal. However, the target emitting the signal is usually on some background (with a temperature of typically, but not mandatory, about 300 K) so that the ultimate limit of the signal discernible by the detector is posed by the fluctuations of the background radiation. In that case we talk about the BLIP device (Background Limited Infrared Photodetector). In real systems, the detection sometimes cannot even reach this limit and is determined by the internal processes within the detector, which in turn are dependent on the particular detection mechanism, as well as on the processes within the amplifier processing further the signal from the detector.

There are a number of parameters quoted in literature as figures of merit of infrared photodetectors. The use of these parameters is dependent on a certain degree on the particular type of the detector.

In this text we adopt Jones' classification, which became standard for the description of photodetectors [2–4]. The main parameters of detectors utilized in this text are collected in Table 1.1.

From the point of view of this work the main parameters are specific detectivity and the detector bandwidth. It is especially convenient to use a synthetic parameter obtained by multiplying the specific detectivity with the bandwidth. Specific detectivity unites sensitivity or quantum efficiency with noise equivalent power. Thus, the D^*f^* product describes all fundamental limitations of a detector, taking into account both its behavior at high operating frequencies and at low signal intensities.

1.1.3 "Ideal" IR Detector

To be able to appraise various photodetector types, we will attempt here to define a general idealized infrared detector, which would, independently on a particular detection mechanism, satisfy all the requirements posed for such a device. The performance of a real detector are then determined by the degree of compliance of its parameters, and thus defined "perfect" device. Of course, one can object that such an approach cannot be objective, since a particular application determines priorities in performance (for instance, for some military applications speed and detectivity are of primary importance, while costs are secondary; for communication purposes one requires large speeds and low prices, etc.). In other words, one cannot define a unique set of characteristics that could describe all existing situations—"an ideal detectors has as many definitions as there are applications" [3].

The approach presented here should be understood in the following manner: our "ideal" device actually sums most of the requirements that could be posed to an IR detector in any application and surrounded by any environment. The purpose of

Table 1.1 Figures of merit of infrared detectors

Name	Notation	Definition	Expression
Responsivity, sensitivity	R	Detector output signal per unit incident radiation power. May be given as voltage responsivity (R_V, S_V, given in V/W), ratio of rms detector output voltage and incident radiation power or current responsivity (R_I, S_I, given in A/W), ratio of rms detector output current and incident radiation power. Usually given either for monochromatic radiation at the signal peak value ($D^*_{\lambda\mathrm{peak}}$) or integral radiation of blackbody source $\left(D^*_{\mathrm{BB}}\right)$, typically at 500 K	$R_v = \frac{V}{P}$ $R_I = \frac{I}{P}$
Quantum efficiency	η	The number of photocarriers (electron-hole pairs) generated per a single photon of incident radiation, i.e., the probability of 1 incident photon generating a charge carrier pair; also the ratio of the current flux and the photon flux. In the systems where all photogenerated carriers contribute to the signal current, equal to the ratio of absorbed and incident power	$\eta = \frac{\mathrm{hc}}{q}\frac{R_I}{\lambda}$
Photoelectric gain	Γ	The number of electrons flowing through the photodetector circuit per an absorbed photon; the ratio of lifetime and transit time	$\Gamma = \tau/T_{\mathrm{tr}}$
Spectral range	λ_{co}	The wavelength above which the photonic detector response quickly starts to drop, corresponds to the photons with an energy equal to the material bandgap, the unit is micrometer	$\lambda_{\mathrm{co}} = 1.239/E_g$ E_g in [eV], λ in [μm]
Noise voltage	V_n	rms signal measured on the detector terminals per a given frequency range in absence of any excitation. The unit is V/Hz$^{1/2}$ or A/Hz$^{1/2}$. Its frequency dependence is called spectral noise or noise spectrum	$U/\Delta f$
Noise current	I_n		$I/\Delta f$
Noise equivalent power	NEP	The incident radiation power generating in the detector an output signal equal to the detector internal noise; corresponds to the minimum detectable flux. Measured in W/Hz$^{1/2}$	$\mathrm{NEP} = \frac{V_n}{R}$
Detectivity	D	Reciprocal value of the noise equivalent power, Hz$^{1/2}$/W	$D = \frac{1}{\mathrm{NEP}}$
Specific detec-tivity, D star	D^*	Signal-to-noise ratio for an incident radiation power of 1 W per an area of 1 cm^2, with noise measured in a frequency interval of 1 Hz (the unit cm Hz$^{1/2}$/W). Usually represented either for monochromatic radiation at the maximum of sensitivity ($D^*\lambda_{\mathrm{peak}}$) or for the integral blackbody radiation $\left(D^*_{\mathrm{BB}}\right)$, usually at 500 K	$D^* = \frac{\sqrt{A\Delta f}}{\mathrm{NEP}}$
D double star	D^{**}	Specific detectivity per an angle θ, independent on the detector field of view (FOV). The unit is cm Hz$^{1/2}$ sterad $^{1/2}$/W	$D^{**} = D^* \sin\theta$

(continued)

Table 1.1 (continued)

Name	Notation	Definition	Expression
Transit time	T_{tr}	Mean time needed for carriers to cross the active area under the influence of the field, inversely proportional to the mobility and the applied field, and directly proportional to the squared path to be crossed, given in seconds	$T_{\mathrm{tr}} = L^2/\mu V$
Carrier lifetime	τ	Mean time of recombination of charge carriers, given in seconds.	$\tau = \Delta n/\Delta g$
Response time	τ_R	The time during which the detector response increases to (1 − 1/e) of its maximum value after the illumination by a Heaviside step of IR radiation, or decreases to 37 % after such signal is switched off, given in seconds. Its reciprocal value is $f_{3\mathrm{dB}}$. Sometimes also defined as the period between the response reaching 10 and 90 % of the full value of the response to a Heaviside step excitation	$V = \frac{V_0}{\sqrt{1+(2\pi f \tau_R)^2}}$
Band-width, cut-off frequency	$f_{3\mathrm{dB}}$	The frequency at which the response to a sinusoidal excitation decreases to 0.707 of its maximum, i.e. its drop is 3 dB, given in Hz	$f^* = 1/4\tau_R$
Detectivity bandwidth	f^*	The operating frequency at which specific detectivity drop is 3 dB, given in Hz	$D^*(f_*^*) \approx 0.707\, D^*_{\max}(f)$
Specific detec-tivity-cutoff frequency product	$D^* f^*$	A derived parameter, the product of specific detectivity and frequency range, describes the behavior of the detector at high frequencies and under low illumination, the unit cm $\mathrm{Hz}^{3/2}/\mathrm{W}$	$D^* f^*$

such an idealization is only to facilitate the comparative analysis of the properties of the existing photodetector types.

The requirements listed in Table 1.2 do not have the same importance, neither taken generally, nor connected to particular practical applications and detector types. The first four are fundamental, while the rest of them depend on concrete implementation.

Some of the parameters listed in Table 1.2 are interconnected (e.g., cooling requirements, costs, reliability and time of preparation for operation, or the number of detector elements and allowed dissipation).

According to the Table, our "ideal" detector would have signal-fluctuation limited operation; it would react momentary when illuminated by IR radiation; its response outside the desired spectral range would be zero; it could not only operate at room temperature, but also on elevated and lower temperatures, and the probability of its failure would be negligible; it would be robust and insensitive to the changes of operating conditions; it would emit neither any harmful radiation nor

Table 1.2 Required performance of IR detectors

	Property	Requirement
1.	Specific detectivity (D^*)	Ideal: limited by signal fluctuations
		Real: BLIP
2.	Responsivity	Maximum
3.	Response time	Shortest possible (ideal: zero)
4.	Spectral response	Maximally selective (full sensitivity in the desired wavelength range, zero outside)
5.	Operating temperature	The largest possible temperature range; uncooled operation
6.	Reliability	As high as possible (ideally: Infinite—no failures)
7.	Environmental influence to the detector	None (minimum sensitivity to operating temperature fluctuations, mechanical shocks, vibrations, radiation, signal or supply overload, etc.)
8.	Detector influence to the environment	None (detector does not emit magnetic field, radiation, heating, static noise, etc.)
9.	Handling	No prior knowledge needed (foolproof)
10.	Electric power dissipation	Minimum (ideal: negligible)
11.	Time of detector preparation for operation	Shortest possible (ideal: zero)
12.	Dimensions and mass	The lowest possible weight and smallest dimensions
13.	Production, operation and maintenance costs	Lowest
14.	Integration with electronic circuitry	Optimal (possibility to furnish standard output)
15.	Number of sensitive elements	Largest possible (maximum resolution within imaging system)

disturbances to surrounding instruments; no special manipulating and maintenance procedures would be necessary, its energy consumption would be negligible; it would be miniature, low-cost, and readily integrable with the existing circuitry, and usable in high-resolution systems. In the next chapter, we consider how different types of real IR detectors fit the described picture.

1.2 Classification of IR Detectors

Numerous physical mechanisms are usable for detection of infrared radiation, and thus there are numerous different detectors. This section lists some more important types of infrared detectors. The key for the presented classification is the basic effect used for the detection of IR radiation.

1.2.1 Thermal Detectors

In this type of detectors material properties are modified by heating, typically independently on the detected radiation wavelength, and this is used to generate electric output or some other type of readout signal. Table 1.3 lists some basic types of thermal IR photodetectors arranged according their detection mechanisms.

For a majority of the existing thermal detectors there is a trade-off between the specific detectivity and the maximum response speed—a detector may be smaller and faster, or larger and more sensitive. There is a frequency-dependent thermodynamical limit of the detectivity of thermal detectors. Instead of that theoretical limit, a more realistic Havens’ limit is used in practical situations [5]. Generally speaking, thermal detectors tend to be at the slower side of the response.

The second limitation is connected with spectral selectivity. Thermal detectors receive radiation in the whole electromagnetic spectrum. They are limited only by the optical properties of the material at the device surface. Their spectral characteristics may be modified e.g., by using a bandpass filter or by properly choosing the material at the device surface.

Some of the quoted performance limitations seriously compromise the applicability of thermal detectors for some key applications, especially those requiring high D^*f^* product.

On the other hand, the majority (although not all—see the last three bolometer-type devices in Table 1.3) thermal detectors do not require cooling not even in far and extreme infrared range. This makes them very convenient for applications where spectral selectivity is not critical. Their integration with processing circuitry is excellent, since thermal detectors are usually fabricated using Si-compatible technologies.

Table 1.3 Basic types of thermal IR detectors

Detector type	Operation principle
Thermocouple	Temperature changes directly cause generation of voltage at a junction of two different materials
Thermopile	An array of a number of thermocouples
Bolometer	Heating changes material resistance
	Metallic (material is a thin metal film)
	Thermistor (material with negative temperate coefficient of resistivity)
	Superconductive (material is superconductor near transition temperature)
	Superinductive (same as before, but inductivity is changed instead of resistance)
	Cryogenic (carbon or semiconductor film at liquid He temperature)
Bimaterial devices	Due to a difference between thermal expansion coefficient, temperature changes the physical shape of the detector, which can be visually observed, i.e., IR radiation is directly converted to a visible signal
Pneumatic detector (Golay cell)	IR irradiated membrane heats a cavity filled with gas and changes its volume
Capacitive microphone with gas fill	As Golay cell, but gas directly absorbs heat and changes microphone capacitance
Pyroelectric detector	Change of electric polarization in some ferroelectric crystals with temperature
Pyromagnetic detector	Change of magnetic properties of material due to heating
Evaporograph	Change of thickness of a thin oil layer condensed on a membrane
Detector with thermo-ionic emission	Change of emission rate from an oxide-coated cathode
Absorption image converter	Temperature variations move semiconductor absorption edge
Liquid crystal detector	Changes of optical properties of liquid crystals with temperature (e.g., reflection coefficient) to achieve direct conversion from IR to visible
Bimaterial detector	Temperature changes cause different thermal expansions of two materials in close contact, resulting in mechanical deformation
Nernst detector	Generation of photovoltage in crossed electric and magnetic field in the presence of thermal gradient

The use of microsystem technologies (MST), especially micromachining, resulted in large advances in thermal detector fabrication, and particularly bolometers and thermopiles. Their reliability and miniaturization are much improved, and power dissipation decreased. MST technologies made possible batch production of two-dimensional arrays of miniature silicon thermal detectors with electronic addressing and a large number of elements.

Table 1.4 Basic types of vacuum photoemissive detectors according to photocathode type

Detector type	Operation principle
Vacuum diodes with metal cathode	Metal is used for photocathode
Devices with negative electron affinity (NEA)	Material with negative electron affinity is used for photocathode, e.g., strongly doped semiconductor coated with cesium

1.2.2 Photoemissive Vacuum Detectors

In this type of detectors, measured optical radiation causes emission of electrons from the surface of a photocathode in a vacuum tube. Electrons may reach directly the anode or first be multiplied on dynodes (photomultiplicators). Historically, these have been the first photodetectors. Table 1.4 presents the basic types of vacuum detection tubes according to the photocathode type.

Regarding its design, a photoemissive detector may be a vacuum tube, photomultiplier (a tube plus dynode converter), some types of camera tubes, image converters, image amplifiers, etc. The main shortcoming of photoemissive tubes is the limited choice of cathode materials, which does not permit their maximum wavelengths to reach the spectral ranges of interest. We mention them for the sake of completeness, as well as because future solutions could potentially overcome this problem.

1.2.3 Detectors with Interaction of Electromagnetic Radiation

The detection process of this type of photodetectors is based on direct interaction of electromagnetic radiation with electromagnetic radiation (e.g., frequency beating, etc.). Table 1.5 shows some basic types of such devices sorted according to the interaction mechanism.

Table 1.5 Detectors with interaction of electromagnetic radiation

Detector type	Operation principle
Optical parametric converter	Optical equivalent of heterodyne effect; the mixing of a coherent signal and a beam from a local oscillator takes place in a nonlinear optical crystal and results in frequency difference (downconversion) or their sum (upconversion)
MOM diode	Metal–metal oxide–metal: electrons tunnel through a thin (of the order of angstroms) oxide layer and IR radiation generates voltage. Extremely fast (100 GHz) and uncooled
Josephson detector	Photoeffect on Cooper pairs in semiconductor within a Josephson junction

Table 1.6 Photonic detectors according to the transition type

Detector type	Operation principle
Intrinsic	Photoexcitation increases the number of free carriers by interband transitions
Extrinsic	Photoexcitation moves carriers between a donor level and the conduction band or an acceptor level and valence band
Free-carrier	Carrier transition takes place between two levels within an energy band (intraband transitions)
Quantum Well IR Photodetector (QWIP)	Intraband photoexcitation of electrons from confined states in the conduction band of quantum wells into the continuum
Quantum Dot IR Photodetector (QDIP)	Intraband photoexcitation of electrons from confined states in the conduction band of quantum dots into the continuum

The detectors from this group are mostly (but not exclusively) limited to the detection of coherent radiation. Their fundamental performance is usually very good; however, their structure is typically rather complex, and this means a higher price and the use mostly in lab environments or in more demanding military applications. Save for a pair of exceptions, detectors based on interaction of electromagnetic waves did not find a wide practical application.

1.2.4 Photonic Detectors

In the case of photonic detectors, IR radiation interacts with charge carriers in semiconductor and generates an electrical signal as a result of a change in the distribution of carriers—i.e., the detection process depends on the photonic nature of the incident radiation. Table 1.6 shows a classification of these detectors according to the nature of the interaction (the type of carrier transitions between bands). Table 1.7 shows the most important types of IR photonic detectors according to the detection mechanisms.

Generally, it may be said that photonic detectors offer the highest detectivity-bandwidth product among all of the presented types of devices. The majority of the photonic detectors are also much faster than thermal detectors, regardless of their specific detectivity. The fabrication technology of semiconductor IR photonic detectors is mature (a consequence of decades of their use in military systems), so that most of the conditions from Table 1.2 are fulfilled. Spectral selectivity can be accurately adjusted for a given application by using bandgap engineering with some of the semiconductors with variable bandgap (e.g., HgCdTe and InAsSb). Arrays with a large number of elements, i.e., with high resolution are fabricated, dissipation per element may be quite low, detectors are robust, compatible with processing circuitry, and reliable (the properties of most of the solid-state devices). At the same time, the structure and design of photonic detectors are the simplest and the related technologies are mature.

Table 1.7 Classification of photonic detectors according to detection mechanism

Detector type	Operation principle
Photoconductor (PC detector)	IR radiation changes electric conductivity of semiconductor material
Photovoltaic detector (detector with p-n junction, PV detector, photovoltaic)	Photoexcitation creates an electron-hole pair to be separated at the p-n junction, thus generating photocurrent. In dependence on the structure type and the junction used it may belong to a number of the existing types: with homojunction or heterojunction, p-i-n photodiode, Camel´s diode (photodiode with modulated barrier), avalanche photodiode (APD), bipolar phototransistor (floating-base), etc.
Schottky detector	Photoexcitation creates an electron-hole pair to be subsequently separated at a metal–semiconductor junction (Schottky junction), thus generating current
MSM detector (metal–semiconductor–metal)	Two Schottky junctions in a structure with interdigital electrode shape
MIS detector	Metal–insulator–semiconductor: photoexcitation generates carriers in the depleted region in a metal–insulator–semiconductor structure
Photoelectromagnetic (PEM) detector	Utilizes photodiffusion, i.e., the appearance of photovoltage in the direction of Ohmic contacts, normally to radiation, as a result of the diffusion of carriers generated by self-photoexcitation in magnetic field (without electric bias). Usually an uncooled device
Dember detector	The appearance of photovoltage in the direction of illumination (normally to the direction of Ohmic contacts) due to the diffusion of carriers generated by self-photoexcitation
Photoemissive Schottky detector (metal–silicide detector)	Utilizes internal photoexcitation at a Schottky barrier, e.g., at PtSi
Bolometer with hot electrons	Photons interact with free electrons in semiconductor and transfer their momentum to them, thus changing their effective temperature
Puttley’s detector	Similar to the bolometer with hot electrons, but transitions take place between the Landau levels in semiconductor in strong magnetic field
Photon drag detector	Photon transfers its momentum to an electron. Very fast, but with low sensitivity
Quantum counter	Illumination excites electron to a higher level, a pump signal excites it to an even higher level. The electron is finally de-excited with an emission of photon detected by a separate detector. Noise of these devices is negligible, but their sensitivity is low
Fluorescent converter	Utilizes stimulated emission of visible radiation in some fluorescent materials previously excited by ultraviolet radiation

The main problem with photonic detectors is that they require cryogenic cooling for their function in MWIR and LWIR infrared ranges in order to minimize thermally induced carrier transitions, and thus avoid unwanted thermal noise.

1.2.5 Choice of Optimum Photodetector

As expected, the overview of the existing photodetector types shows that none can satisfy in full the requirements listed in Table 1.2. Therefore, we proceed by the method of elimination.

From the point of view of MWIR and LWIR detection the today's photoemissive devices must be rejected first because of the physical limitations posed by their short cutoff wavelength, since it is difficult to imagine a way to overcome this fundamental limitation. On the other hand, we also reject the much more attractive detectors with interaction of electromagnetic waves, because today these are much more exotic and complex lab systems than practical detective devices—although we should mention that their further development could hypothetically bring to an appearance of novel or modified detectors with characteristics much closer to our imaginary ideal photodetector than today's devices. Some of the detection systems based on the interaction of radiation have outstanding performance (e.g., heterodyne detectors) but one of their subassemblies is conventional detector, so that by definition they cannot belong to the group of choice.

A comparison of the remaining two classes, photonic and thermal detectors, shows that their performance is similar from the point of view of the criteria presented in Table 1.2. The main advantage of photonic detectors is their response speed, and in the case of thermal detectors their room-temperature operation and relatively lower price. This means that each of these detector types can find a share of market where its will dominate.

From the point of view of the D^*f^* parameter, not taking into account the other criteria, photonic detectors appear as a better choice. To be sure, microsystem technologies enable a simultaneous increase of both the speed and the detectivity of thermal detectors, but only up to a certain point. According to [6], not only that the maximum specific detectivity and the bandwidth of these devices are inversely proportional, but also the maximum attainable specific detectivity at room temperature is limited by integral radiation of the background, not only by its part covered by spectral sensitivity range of the detector, as is the case with photonic devices. The conclusion is that a photonic detector without cooling requirements would have a chance to approach nearer to the idealized device described in Sect. 1.3.

This work is dedicated to an analysis of the possible improvements of the performance of photonic detectors with a goal to use them at room temperature and to approach them as much as possible to the previously described idealized IR detector element. The presentation is limited to photoconductive and photovoltaic intrinsic detectors, although for the most part it can be generalized to any type of photonic detectors. Often it can be even directly applied to such devices, without any modifications. A number of the presented approaches can actually be utilized for a majority of other detector types.

1.3 On Specific Detectivity of Photonic Infrared Detectors

In this section, we consider various possibilities to maximize the $D^{*}f^{*}$ product of photonic IR detectors. To this purpose we proceed to analyze separately specific detectivity and response time.

We start from the phenomenological expression for the current of a photoconductive detector, described in literature as the fundamental equation of photoconductivity [7]. This expression is valid both for intrinsic and extrinsic detectors in stationary state and is given as

$$I_f = \eta q A \Phi \Gamma. \tag{1.1}$$

This is actually the increase of the dark current due to illumination and is a dc value (valid for $f = 0$ Hz). Here η denotes the quantum efficiency of the detector, ϕ is the incident photon flux density, A is the detector active area. The factor Γ denotes the photoelectric gain or photogain (the ratio between the number of electrons flowing through the electric circuit and the number of absorbed photons). The fundamental equation of photoconductivity is also valid without changes for the short circuit current of a photovoltaic detector (photodiode operating in photoconductive mode). In that mode of operation $\Gamma \approx 1$ in most of the cases.

Using the fundamental equation of photoconductivity and the definition of the current responsivity given in Table 1.1

$$R_{I\lambda}(\lambda) = \frac{q\eta\lambda\Gamma}{\mathrm{hc}}, \tag{1.2}$$

we obtain for specific detectivity, again according to the definition

$$D_{\lambda}^{*} = \frac{q}{\mathrm{hc}}\eta\lambda\sqrt{A}\Gamma\left(\frac{I_n}{\sqrt{\Delta f}}\right)^{-1}. \tag{1.3}$$

In the case when the surface of the detector receiving the incident optical flux is not equal to the detector active area, i.e., when some kind of concentrator is used to collect radiation from a larger area and direct it to the active area, a factor of optical concentration [8] may be formally introduced into the expression for the specific detectivity. This factor is equal to the square root of the ratio between the optical and active ("electrical") detector area.

The expression (1.3) is the basic relation for our analysis of possibilities for the optimization of photonic detector performance. From this expression, it can be seen that specific detectivity increases with wavelength if noise current is mostly independent on incident radiation wavelength, and if quantum efficiency is approximately constant within the spectral range under consideration. This is valid until the cutoff wavelength is reached. This wavelength corresponds to the band edge of the semiconductor used. At this value the absorption coefficient starts to drop quickly, decreasing quantum efficiency, and thus the specific detectivity as well. This is the reason why the ideal spectral dependences of responsivity and detectivity of a photonic detector have the typical "saw" profile: their value first linearly increases with wavelength, and then quickly drops near the cutoff. In reality, these spectral characteristics are usually more or less flattened near the peak [9] It is commonly assumed that the optimum wavelength for the detector operation is near its cutoff wavelength.

Specific detectivity is proportional to photoelectric gain. This is why it is convenient to use e.g., avalanche devices, where the photogain is large. However, the photogain increase must not compromise other factors, for instance response time or noise level.

The third factor to be maximized in order to improve specific detectivity is quantum efficiency. This parameter depends on a number of factors. In further text, we shortly consider the expressions for quantum efficiency for the case of photovoltaic and photoconductive IR detectors.

An expression for the quantum efficiency of photoconductors, taking into account reflection coefficients from both the front and the back side of detector, as well as surface recombination rates on both sides (interference effects are neglected), is [10]

$$\eta = \frac{1 - R_1}{(\alpha^2 L_a^2 - 1)(1 - R_1 R_2 e^{-2\alpha d})} \left\{ F(\alpha, S_1, S_2) - R_2 e^{-2\alpha d} F(-\alpha, S_1, S_2) \right\}, \tag{1.4}$$

where $L_a = \sqrt{D_a \tau}$ is the ambipolar diffusion length, D_a is the ambipolar diffusion coefficient, R_1 and R_2 are the reflection coefficients from the front and the back side of detector, S_1 and S_2 are the corresponding surface recombination rates, α is the absorption coefficient of the active area material, d is the path of the radiation through the active region without taking into account reflections on the front and the back side (i.e., the physical thickness of the active region). The factor F is

$$\begin{aligned} F(\alpha, S_1, S_2) = {} & \frac{\alpha L_a \left\{ \left(\alpha L_a - \frac{L_a S_2}{D_a} \right) \left[\frac{L_a S_1}{D_a} \left(1 - \mathrm{ch}\left(\frac{d}{L_a} \right) \right) - \mathrm{sh}\left(\frac{d}{L_a} \right) \right] e^{-\alpha d} + F_1 \right\}}{\frac{L_a (S_1 + S_2)}{D_a} (S_1 + S_2) \mathrm{ch}\left(\frac{d}{L_a} \right) + \left(1 + \frac{L_a^2 S_1 S_2}{D_a^2} \right) \mathrm{sh}\left(\frac{d}{L_a} \right)} \\ & - \left(1 - e^{-\alpha d} \right), \end{aligned} \tag{1.5}$$

$$F_1 = \left(\alpha L_a - \frac{L_a S_1}{D_a} \right) \left[\mathrm{sh}\left(\frac{d}{L_a} \right) + \frac{L_a S_2}{L_a} \left(\mathrm{ch}\left(\frac{d}{L_a} \right) - 1 \right) \right]. \tag{1.6}$$

For the case $S_1 = S_2 = S_0$ the above expression becomes

$$\eta = \frac{(1 - R_1)\left(1 + R_2 e^{-\alpha d}\right)\left(1 - e^{-\alpha d}\right)}{\left(1 - R_1 R_2 e^{-2\alpha d}\right)\left(\alpha^2 L_a^2 - 1\right)} \left[\alpha L_a \frac{\alpha L_a + \frac{S_0 L_a}{D_a} c\mathrm{th}\left(\frac{\alpha d}{2}\right)}{1 + \frac{S_0 L_a}{D_a} c\mathrm{th}\left(\frac{\tau}{2L_a}\right)} - 1 \right]. \tag{1.7}$$

Further simplification is obtained for $S_0 = 0$

$$\eta = \frac{(1 - R_1)\left(1 - e^{-\alpha d}\right)\left(1 + R_2 e^{-\alpha d}\right)}{1 - R_1 R_2 e^{-2\alpha d}}. \tag{1.8}$$

If $R_1 = 0$, $R_2 = 1$ the expression becomes

$$\eta = 1 - e^{-2\alpha d}. \tag{1.9}$$

If the absorption coefficient times detector thickness product is sufficiently larger than 1, quantum efficiency becomes approximately 1.

A general expression for the quantum efficiency of photovoltaic detectors taking into account surface recombination rates and reflection coefficients on both of its surfaces was proposed by Djurić and Jakšić [11]. Quantum efficiency within the n-region of the detector is

$$\eta_n = \frac{1 - R_1}{1 - R_1 R_2 e^{-2\alpha H_d}} \left[F_n(\alpha) - R_2 e^{-2\alpha H_d} \cdot F_n(-\alpha)\right], \tag{1.10}$$

where

$$F_n(\alpha) = \frac{\alpha L_n}{\alpha^2 L_n^2 - 1} e^{-\alpha(x_j + w)} \left(\alpha L_n - \frac{\frac{d\left[\frac{S_n L_n}{D_n}\mathrm{sh}\left(\frac{H'}{L_n}\right) + \mathrm{ch}\left(\frac{H'}{L_n}\right)\right]}{d(H'/L_n)} + \left(\alpha L_n - \frac{S_n L_n}{D_n}\right) e^{-\alpha H'}}{\frac{S_n L_n}{D_n}\,\mathrm{sh}\left(\frac{H'}{L_n}\right) + \mathrm{ch}\left(\frac{H'}{L_n}\right)} \right) \tag{1.11}$$

where x_j denotes the position of the p-n junction calculated from the n-side, w is the width of the depleted zone, H' is the width of the p-zone, H_d is total detector thickness, L_n is the electron diffusion length, D_n is the electron diffusion constant.

The quantum efficiency of depleted region is

$$\eta_w = \frac{1 - R_1}{1 - R_1 R_2 e^{-2\alpha H}} \left(F_w - R_2 e^{-2\alpha H} F_w\right), \tag{1.12}$$

Here the factor F becomes

$$F_w = e^{-\alpha x_j}\left(1 - e^{-\alpha w}\right). \tag{1.13}$$

The p-region has a quantum efficiency of

$$\eta_p = \frac{1 - R_1}{1 - R_1 R_2 e^{-2\alpha H_d}} \left[F_p(\alpha) - R_2 e^{-2\alpha H_d} F_p(-\alpha) \right], \tag{1.14}$$

where the factor F_p is

$$F_p(\alpha) = \frac{\alpha L_p}{\alpha^2 L_p^2 - 1} \left(\frac{\left(\frac{S_p L_p}{D_p} + \alpha L_p \right) - e^{-\alpha x_j} \frac{d\left(\frac{S_p L_p}{D_p} \mathrm{sh}(x_j/L_p) + \mathrm{ch}(x_j/L_p) \right)}{d(x_j/L_p)}}{\frac{S_p L_p}{D_p} \mathrm{sh}\left(\frac{x_j}{L_p} \right) + \mathrm{ch}\left(\frac{x_j}{L_p} \right)} - \alpha L_p e^{-\alpha x_j} \right) \tag{1.15}$$

The total quantum efficiency of a photovoltaic detector is

$$\eta_{\mathrm{tot}} = \frac{\eta_p + \eta_n + \eta_w}{1 - R_1}. \tag{1.16}$$

In practical situations the largest influence is that of the depletion region. This means that the form of the dependence is basically similar to the case of photoconductive detector.

By considering the above expressions, we arrive at the obvious conclusion that quantum efficiency can be maximized by decreasing the front side reflection coefficient and increasing it on the back side, as well as by maximizing the absorption coefficient.

The last remaining factor within the expression for specific detectivity is noise current that has to be minimized. This factor is considered in the next two chapters.

The second term to optimize within the $D^* f^*$ product is the response time. This factor is very complex and in a photonic detector it is determined, among others, by the time of carrier diffusion to the depletion region, the time of transit across the active region, the carrier sweepout time (in photoconductors), the detector RC constant, and by the RC constant of external circuitry [7]. The response time is fundamentally limited by the carrier transit time [8], and this time is inversely proportional to the thickness of the active region.

According to the definitions given in Table 1.1 the transit time can be maximized by minimizing the active region thickness, by increasing bias to the maximum in the drift velocity dependence on electric field and by choosing semiconductor material with larger carrier mobility. In Part III the response time is considered in more detail, connected with specific types of photodetectors.

Finally, as an illustration, a frequency dependence of the specific detectivity-bandwidth product for two different types of infrared detectors is given, one of them photonic and the other thermal device. One of the detectors is a HgCdTe

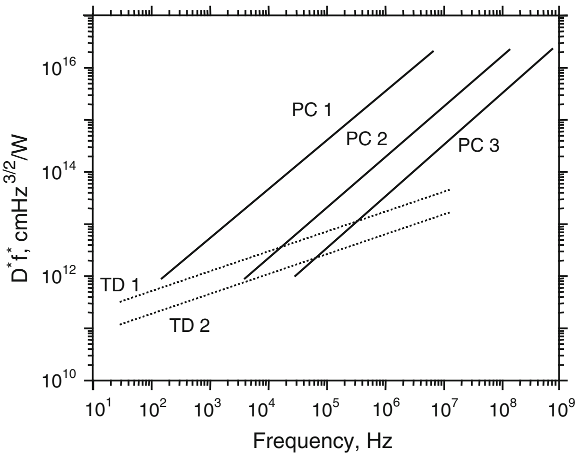


Fig. 1.1 Specific detectivity-cutoff frequency product versus operating frequency for a mercury cadmium telluride photoconductor (denoted as PC) and a pyroelectric detector (TD). PC 1 is a photoconductor at 200 K, PC 2 at 250 K, PC 3 at 300 K. TD 1 is the theoretical performance maximum for a pyroelectric detector, TD 2 are values achieved in laboratory

photoconductor, and the other is a pyroelectric detector. It can be seen that the D^*f^* product performance of photonic detectors largely exceeds those of thermal detectors under similar operating conditions (Fig. 1.1).

1.4 Recombination Mechanisms in Direct Narrow-Bandgap Semiconductor

Generation-recombination (g-r) of charge carriers determines the basic behavior of semiconductor devices. At the same time, it defines concentrations of both carrier types and the response to optical signal. Generation-recombination processes influence the value of specific detectivity by determining the quantum efficiency and at the same time their stochastic fluctuations cause a noise component that is in most cases prevailing over other mechanisms. At the same time, g-r processes determine the time response of the devices. Accordingly, the lifetime of photogenerated charge carriers is one of the most important parameters of detector material. It may be defined as a carrier concentration change caused by the net generation-recombination rate [8]

$$\tau_n = \frac{\mathrm{d}n}{\mathrm{d}g_n}, \quad \tau_p = \frac{\mathrm{d}p}{\mathrm{d}g_p} \tag{1.17}$$

for electrons (n) and holes (p), respectively. Here dn denotes the electron concentration increment caused by photogeneration, and d$g_{n,p}$ is the increment of the net g-r rate (the latter being defined as the difference between the generation and the recombination rate).

There is a large number of bulk and surface mechanisms determining net generation-recombination rate in semiconductor devices. One of the usual approaches for taking them into account is to include surface recombination processes directly into the expressions for quantum efficiency, while bulk processes are calculated separately (e.g., [12, 13]).

The basic bulk g-r processes in direct narrow-bandgap semiconductors are radiative, Auger, and Shockley-Read [8]. Of these three, Auger processes and radiative transitions are fundamental (determined by the physical nature of the processes occurring in semiconductor materials, i.e., by the existence of bandgap,) while Shockley-Read processes are nonfundamental and may be minimized by improving technological processes for fabrication of semiconductors. Although there are certain discrepancies in literature, a generally accepted picture is that in sufficiently pure narrow-bandgap semiconductor single crystals at temperatures in the range (200–300) K Auger recombination prevails, together with its inverse impact ionization (Auger generation) (e.g., [14–16]). If the temperature of the active region is low enough or the operating wavelength short enough (visible or ultraviolet), detector performance is limited by radiative processes. In materials with larger concentrations of traps Shockley-Read processes are prevailing.

The further text shortly considers all of the three main bulk generation-recombination mechanisms in direct narrow-bandgap semiconductors. A unified g-r term is presented to be utilized in modeling and optimization of photonic IR detectors. In addition to that, some useful approximations are given that may significantly speed up and simplify modeling.

1.4.1 Radiative Processes

Radiative processes include generation of free carriers due to thermal radiation and the inverse mechanism, the recombination of a free electron and a free hole with a simultaneous emission of a light quantum. They are interband processes, directly determined by the semiconductor band structure, and thus represent an intrinsic or fundamental mechanism. A theory of radiative g-r processes was first given by van Roosbroeck and Shockley [17]. Figure 1.2 shows a schematic presentation of radiative recombination and generation.

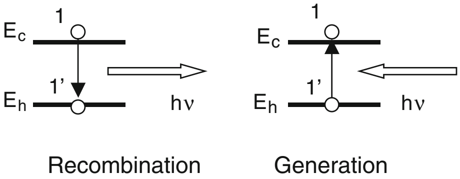


Fig. 1.2 Schematic presentation of radiative recombination and generation

In thermal equilibrium the radiative recombination rate is equal to the generation rate due to thermal radiation for each frequency interval dv. If $W(v)$ is the probability for a photon with a frequency ν to be absorbed in a unit time, and $\rho(v)$ energy density of the photon in a given volume of semiconductor crystal per dv, the radiative generation rate in thermal equilibrium is

$$G_{r0} = \int_{v_{\mathrm{co}}}^{\infty} W(v) \frac{\rho(v)}{\mathrm{h}\nu} \mathrm{d}v, \tag{1.18}$$

where ν_{co} is the cutoff frequency, i.e., the frequency corresponding to the cutoff wavelength. The probability of photon absorption is [17]

$$W(v) = \alpha(v) \frac{c}{\sqrt{\varepsilon(v)}}, \tag{1.19}$$

where $\varepsilon(v)^{1/2}$ is the real part of semiconductor refractive index (square root of the real part of its dielectric permittivity).

Photon energy density is given by the Planck law

$$\rho(v) = \frac{8\pi v^2}{c^3/\varepsilon^{3/2}} \frac{\mathrm{h}\nu}{\exp(\mathrm{h}\nu/k_bT) - 1}. \tag{1.20}$$

Thus spontaneous radiative generation rate is

$$G_{r0} = \frac{8\pi}{c^2} \int_{v_{\mathrm{co}}}^{\infty} \frac{\varepsilon(v)\alpha(v)v^2 \mathrm{d}v}{\exp(\mathrm{h}\nu/k_bT) - 1}. \tag{1.21}$$

Instead of using the dispersion of dielectric permittivity in (1.21), the approximation $\varepsilon(v) \approx \varepsilon_\infty$ is often applied in literature (ε_∞ is the high-frequency value of effective dielectric permittivity) [13], so that we adopt the same approach in this text.

The semiconductor absorption coefficient $\alpha(v)$ can be determined from experimental values, numerically calculated according to accurate theoretical dependences or analytically determined using approximations.

Hall's expression for spontaneous radiative recombination rate, obtained using Bardeen's approximation for the absorption coefficient of narrow-bandgap semiconductors $\alpha \sim \left(E - E_g\right)^{1/2}$ [18] is often used to calculate the generation rate G_{r0} [19, 20]. A generalization of Hall's expression is given as [29, 351]

$$\begin{aligned} G_{r0} = n_i^2 \times 5.8 \\ \times 10^{-13} \sqrt{\varepsilon_\infty} \left(\frac{m_0}{m_n^* + m_p^*} \right)^{3/2} \left(1 + \frac{m_0}{m_n^*} \right) \left(\frac{300}{T} \right) \left(E_g^2 + 3E_g k_b T + 3.75 k_b^2 T^2 \right), \end{aligned} \tag{1.22}$$

where energy is in (eV), intrinsic concentration n_i in cm^{-3}, and G_{r0} is obtained in cm^{-3} s^{-1}.

Hall's expression is often used even for narrow-bandgap semiconductors with nonparabolic bands and with an arbitrary degeneration [19, 21, 22]. Most of the calculations in this work were performed using the accurate model of absorption coefficient, based on Kane's model with taking into account the Urbach band tailing [23]. For some calculations an empirical approximation for HgCdTe absorption coefficient was used [24]

$$\alpha = (1480x + 0.26T + 90)e^{3.915\left(E-E_g\right)^{1/3}}\left[\mathrm{th}\beta\left(E - E_g\right) + 1\right], \tag{1.23}$$

where energy is in eV, and α in cm^{-1}. E_g was calculated according to the expression of Hansen et al. [25]. The parameter β is

$$\beta\left\lfloor \mathrm{eV}^{-1}\right\rfloor = 120\mathrm{th}(10x - 1.5). \tag{1.24}$$

High-frequency dielectric permittivity was calculated according to [26].

In nonequilibrium generation rate G_r is proportional to np, and in thermal equilibrium ($np = n_i^2$) equal to G_{r0}

$$G_r = \frac{np}{n_i^2} G_{r0}. \tag{1.25}$$

For small excess concentrations (Δn, Δp) ≪ (n, p) the deviation of the generation rate from the equilibrium value is

$$\Delta G_r = \frac{\Delta n p_0 + \Delta p n_0 + \Delta n \Delta p}{n_0 p_0} G_{r0}. \tag{1.26}$$

If $\Delta n = \Delta p$, radiative lifetime is according to the definition

$$\tau_r = \frac{n_i^2}{G_{r0}(n_0 + p_0 + \Delta n)}. \tag{1.27}$$

Maximum radiative lifetime is obtained for intrinsic semiconductor (intrinsic radiative lifetime)

$$\tau_{ri} = \frac{n_i}{2G_r}. \tag{1.28}$$

Figure 1.3 shows the dependence of radiative lifetime on inverse temperature for a mercury cadmium telluride sample with a composition $x = 0.195$ for various electron concentrations.

The calculation was performed by direct numerical integration of (1.21) and using (1.27). The calculation did not take into account bandgap narrowing due to

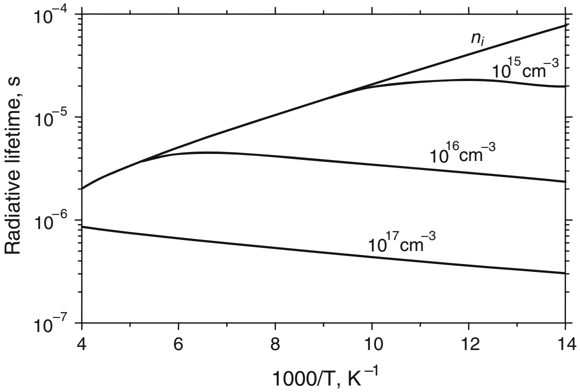


Fig. 1.3 Radiative lifetime versus reciprocal temperature for a mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) sample $x = 0.195$ for different carrier concentrations

free carrier interaction with other free carriers and with holes. Radiative lifetime is shorter if bandgap filling is not taken into account, especially for lower temperatures and higher concentrations.

All expressions on radiative recombination presented in this text until this point denoted microscopic radiative lifetime, i.e., single acts of radiative recombination (a "measure" of the efficiency of photons escaping from the detector bulk). However, a photon emitted within a semiconductor sample during recombination may be absorbed again, which effectively increases total radiative lifetime [27]—the phenomenon of reabsorption or photon recycling [28, 29].

When considering photon recycling processes one can always neglect the component of radiative recombination caused by the photons, which do not leave the sample. This is because reabsorption processes are very fast, so that they do not contribute to the lifetime within the photodetector bandwidth. If the real part of the refractive index of semiconductor is n_{ref}, α its absorption coefficient and c velocity of light, the photon lifetime will be $(\alpha v_g)^{-1} = n_{\mathrm{ref}}/\alpha c$, i.e., in real situations $(10^{-13}\text{–}10^{-12})$ s, several orders of magnitude smaller than carrier recombination lifetime. This means that reabsorption acts do not influence the noise level within the detector and only total reabsorption radiative lifetime is important for its performance.

Let us now calculate radiative lifetime including reabsorption. Radiative recombination rate per unit frequency for the case of reabsorption can be represented as Eq. (1.25) multiplied by P_E, probability for a photon to leave the sample after being generated in electron-hole recombination

$$R_{\mathrm{rad}} = R_r \frac{np}{n_i^2} P_E = W(v) f(v) \rho_v. \tag{1.29}$$

The parameter R_r was calculated using the principle of detailed balance (the number of recombination acts followed by the emission of a photon with a frequency within the interval v and $v + \mathrm{d}v$ is equal to the number of electron-hole pairs generated by equilibrium thermal radiation in the same interval). Here, $f(v)$ denotes the

Bose-Einstein distribution function, $\rho\nu$ is density of states, while $W(\nu) = \alpha v_g = \alpha c/n_{\mathrm{ref}}$ is probability of photon absorption per unit time.

On the other hand, radiative generation is calculated as

$$G_{\mathrm{rad}} = \frac{c\, q_0 A}{4 n_{\mathrm{ref}}^3 V} P_A, \tag{1.30}$$

where P_A is the probability of absorption of incoming photons, A is the area and V the volume of the detector ($V = Ad$).

According to the principle of detailed balance, G_{rad} and R_{rad} must be equal in thermodynamical equilibrium. From this condition, we may determine the equilibrium probability of the emission of a photon generated by a radiative recombination act, P_{e0}. For small deviations from equilibrium we assume $P_E \approx P_{E0}$. In this manner, we determine the difference between radiative generation and recombination, i.e., the net rate of radiative processes

$$g_{\mathrm{rad}} = \left(\frac{np}{n_i^2} - 1\right) \frac{c}{4 n_{\mathrm{ref}}^3 d} q_0(\nu) P_A(\nu). \tag{1.31}$$

Photon density per unit frequency $q_0(\nu)$ for the case of thermal equilibrium is given as

$$q_0(\nu) = \frac{8\pi \nu^2 n_{\mathrm{ref}}^3}{c^3 \left(e^{\mathrm{h}\nu/kT} - 1\right)}, \tag{1.32}$$

if the dimensions of the detector are large enough compared to the operating wavelength.

After performing integration over frequency we obtain the following expression for the net rate

$$g_{\mathrm{rad}} = \left(\frac{np}{n_i^2} - 1\right) \frac{c}{d} \int\limits_{E_g/h}^{+\infty} \frac{1}{4 n_{\mathrm{ref}}^3} \frac{8\pi \nu^2 n_{\mathrm{ref}}^3}{c^3 \left(e^{\mathrm{h}\nu/kT} - 1\right)} P_A \mathrm{d}\nu. \tag{1.33}$$

To calculate this expression we must know the absorption probability P_A. To this purpose we use the expression for flux taking into account reflection on both sides (at the same time neglecting interference effects)

$$g(y) = \frac{(1 - R_1)\Phi}{1 - R_1 R_2 \exp(-2\alpha d)} \left[\exp(-\alpha y) + R_2 \exp(-2\alpha d + \alpha y)\right], \tag{1.34}$$

so that we obtain

$$\begin{aligned} P_A &= \frac{1}{q_0}\int\limits_0^d \left[\frac{\alpha(1-R_1)q_0}{1-R_1R_2e^{-\alpha d}}\left[e^{-\alpha y}+R_2e^{-2\alpha d+\alpha y}\right]+\frac{\alpha(1-R_2)q_0}{1-R_1R_2e^{-\alpha d}}\left[e^{-\alpha y}+R_1e^{-2\alpha d+\alpha y}\right]\right]\mathrm{d}y \\ &= \frac{1-e^{-\alpha d}}{1-R_1R_2e^{-\alpha d}}\left[2-(R_1+R_2)\left(1-e^{-\alpha d}\right)-2R_1R_2e^{-\alpha d}\right] \end{aligned} \tag{1.35}$$

Finally we use the definition (1.17) to obtain the expression for radiative lifetime

$$\frac{1}{\tau_R}=\frac{g_{\mathrm{rad}}}{\Delta n}=\frac{n_0+p_0}{n_i^2}\frac{c}{d}\int\limits_{E_g/h}^{+\infty}\frac{1}{4n_{\mathrm{ref}}^3}\frac{8\pi v^2 n_{\mathrm{ref}}^3}{c^3\left(e^{\mathrm{hv}/kT}-1\right)}P_A(\mathrm{v})\mathrm{d}v. \tag{1.36}$$

According to this, in intrinsic semiconductor ($n_0 = p_0 = n_i^2$) the radiative lifetime τ_{ri} becomes

$$\tau_{ri}=\frac{\mathrm{d}n_i}{2\int_{E_{g/h}}^{+\infty}\frac{2\pi v^2}{c^2\left(e^{\mathrm{hv}/kT}-1\right)}\frac{1-e^{-\alpha d}}{1-R_1R_2e^{-\alpha d}}\left[2-(R_1+R_2)(1-e^{-\alpha d})-2R_1R_2e^{-\alpha d}\right]\mathrm{d}v}. \tag{1.37}$$

1.4.2 Auger Processes

Auger mechanism denotes a group of nonradiative interband generation and recombination processes. All of these processes are fundamental (determined by semiconductor structure). They are three-particle processes—the excess energy appearing as a result of the interaction of two carriers is transferred to a third one and its energy shifts to a higher energy level within the same band.

Auger processes encompass Auger recombination and its opposite impact ionization. Beattie defined ten basic Auger processes in material with a single conduction band and with a heavy holes and light holes band. In the processes of nonradiative interband recombination phonon states, localized states and impurity levels may take part. Landsberg described 70 such secondary Auger processes [30].

In narrow-bandgap direct semiconductor of the $Hg_{1-x}Cd_xTe$ and InSb type the dominant processes are Auger 1 (CCCH) and Auger 7 (CHHL) (Fig. 1.4). Here, the letters in the notation denote the bands containing the carrier taking part in the Auger process. The first two letters denote the initial state, and the second two the final one. *C* means the conduction band, *H* is the heavy holes band, and *L* is the light holes band. *S* means the spin split-off band.

The threshold energy for the Auger 1 and Auger 7 processes is the lowest, and their total density of states, and thus their probability of occurrence, the largest. In the CCCH Coulomb interaction appears between two electrons in conduction band (states 1 and 2). Due to this, an electron crosses into the heavy holes band and

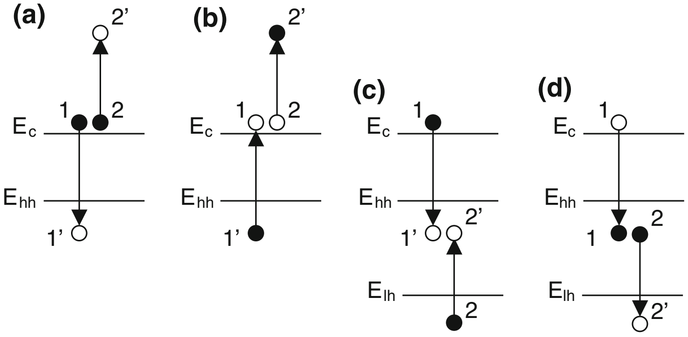


Fig. 1.4 Schematic presentation of Auger 1 (CCCH) and Auger 7 (CHHL) process. Numbers without "prime" represent the state before the transition, and primed numbers are the state afterward. Electrons are *solid black*, *holes empty circles*. **a** CCCH recombination; **b** impact ionization inverse to it; **c** CHHL; **d** impact ionization inverse to it

recombines (state 1′), while the second electron is excited to a higher level within the conduction band (state 2′). The process ends by the electron 2′ emitting a phonon and returning to its bottom level. In the CHHL process two holes from the heavy hole band take part. An electron transfers from the conduction band to the heavy hole band and recombines with one of the holes. A second electron arrives from the light hole band and recombines with the heavy hole.

CCCH mechanism prevails in n-type material, and CHHL dominates in p-type narrow-band semiconductors.

The basic theory of CCCH processes and Auger recombination in semiconductors was given in 1959 by Beattie and Landsberg in their seminal paper [31], while the CHHL process was examined by Beattie and Smith [32]. Takeshima showed that Auger S process (CHHS) also becomes marked in materials with a bandgap comparable to spin split-off band energy [33, 34]. Sugimura proposed a generalization of the Auger S process [35, 36].

Also of interest for the Auger recombination in narrow-bandgap semiconductors, especially in HgCdTe are the papers of Gel'mont et al. [15, 16, 37], of Haug [38, 39], Gerhardts et al. [40], Casselman et al. [41] and Bertazzi [42], as well as the books by Landsberg [43] and Blakemore [44].

The overlap integral, one of the basic parameters of Auger processes, was approximately calculated by Beattie and Landsberg based on the Kronig-Penney model [45], as well as by Antončik and Landsberg [46]. The calculations of the overlap integral were critically reexamined in a paper by Burt et al. [47]. For approximate calculations, today, one often uses empirical values of overlap integrals [8], typically in the range 0.1–0.3, which gives a correct order of magnitude of the Auger lifetime.

One of the problems on Auger recombination is connected with the expansion of the analytical expression of Beattie and Landsberg to degenerate semiconductor. A body of papers is dedicated to accurate calculation of Auger lifetime and overlap integrals for various materials and different degrees of degeneration [48, 49].

In order to present some basic expressions for the Auger lifetime let us denote Auger generation (impact ionization) rate as G_A and Auger recombination rate as R_A. We consider first the Auger 1 process. Since two electrons and one hole take part in it, the recombination rate will be proportional to n^2p

$$R_{A1} = \frac{n^2 p}{n_i^3} R_{A1}^i. \tag{1.38}$$

Auger generation is proportional to electron concentration

$$G_{A1} = \frac{n}{n_i} G_{A1}^i. \tag{1.39}$$

In a thermodynamical equilibrium according to the principle of detailed balance $R_{A1} = G_{A1}$, so that $R_{A1}^i = G_{A1}^i$.

The net recombination rate of Auger 1 process is obtained by subtracting expressions (1.38) and (1.39)

$$r_{A1} = R_{A1} - G_{A1} = G_{A1}^i \frac{np - n_i^2}{n_i^2} \frac{n}{n_i}. \tag{1.40}$$

According to the definition of lifetime (1.17)

$$\tau_{A1} = \frac{\Delta n}{r_{A1}} = \frac{2 n_i^2 \tau_{A1}^i}{(p_0 + n_0)(n_0 + \Delta n)}, \tag{1.41}$$

i.e.,

$$\tau_{A1} = \frac{n_i^3}{G_{A1}^i (p_0 + n_0)(n_0 + \Delta n)}. \tag{1.42}$$

Accordingly, Auger lifetime in intrinsic semiconductor is

$$\tau_{A1}^i = \frac{n_i}{2 G_{A1}^i}. \tag{1.43}$$

For Auger 7 processes recombination rate is proportional to p^2n, thus

$$R_{A7} = \frac{p^2 n}{n_i^3} R_{A7}^i. \tag{1.44}$$

By a procedure equivalent to (1.38)–(1.43) we obtain for Auger 7 process

$$r_{A7} = G^i_{A7} \frac{np - n_i^2}{n_i^2} \frac{p}{p_0}, \tag{1.45}$$

$$\tau_{A7} = \frac{n_i^3}{G^i_{A7}(p_0 + n_0)(n_0 + \Delta n)}. \tag{1.46}$$

According to the approach of Beattie and Landsberg [31], the intrinsic Auger lifetime is calculated by perturbation method. The perturbation operator of Auger mechanism, i.e., of Coulomb interaction of two electrons, is calculated by subtracting Hartree-Fock single-electron Hamiltonian from the “complete” Hamiltonian of the system (and actually the simplest Hamiltonian that still “sees” the Coulomb interaction of the Auger process). It has the form of a screened Coulomb potential

$$V(r) = \frac{q^2}{4\pi\varepsilon_0\varepsilon_r r}\exp\left(-\frac{\lambda_D}{r}\right), \tag{1.47}$$

where r denotes position, ε is relative dielectric permittivity, λ_D is Debye length of the screened potential.

The square of the matrix element corresponding to the above perturbation operator gives T_{if}, the probability of Auger 1 transitions [31]

$$T_{if} = \frac{2q^4t^2}{V^2\hbar^2\varepsilon_0^2\varepsilon_r^2}\left\{\left[\frac{F_1F_2}{\lambda_D^2 + k_g^2} - \frac{F_3F_4}{\lambda_D^2 + k_h^2}\right]^2 + \left[\frac{F_1F_2}{\lambda_D^2 + k_g^2}\right]^2 + \left[\frac{F_3F_4}{\lambda_D^2 + k_h^2}\right]^2\right\} \times \frac{1 - \cos\omega_{fi}t}{(\omega_{fi}t)^2}\delta\left(\vec{k}_1 + \vec{k}_2 - \vec{k}_{1'} - \vec{k}_{2'}\right). \tag{1.48}$$

V is the volume, t time, $\omega_{fi} = |E_f - E_i|/\hbar$ (angular frequency corresponding to the difference between the final f and the initial i energy state), $\vec{k}_j$ are wavevectors of charge carriers in the j-th state j (j = 1, 2, 1′, 2′), $\vec{k}_g = \vec{k}_1 - \vec{k}_{1'}$, $\vec{k}_h = \vec{k}_2 - \vec{k}_{2'}$, while F_1, F_2, F_3 and F_4 are the overlap integrals of the modulating parts of Bloch wave functions for the states 1–1′, 2–2′, 1–2′ *i* 2–1′, respectively.

The Auger lifetime is calculated according to the definition (1.17), so that

$$\tau_A = \frac{\Delta n}{\Delta G_A} = \frac{\Delta n V t}{P_A}. \tag{1.49}$$

The net probability of the Auger 1 transition P_A is obtained by integrating the expression for T_{if}:

$$P_A = \int T_{if}(w_1w_{1'}w_2w_{2'} - (1 - w_1)(1 - w_{1'})(1 - w_2)(1 - w_{2'}))\mathrm{d}k\mathrm{d}r, \tag{1.50}$$

where w_j is the probability of the occupation of the state j by a corresponding particle (electron or hole), and $1 - w_j$ is the probability for the vacancy of this state.

Thus, the squared modulus of the Coulomb interaction matrix element is multiplied by the probabilities of the occupancy of the initial and the vacancy of the final state and integrated over wavevectors of all of these states. At that, the laws of energy and quasimomentum conservation must be satisfied.

The above expression for Auger transition probability is generally applicable for all degrees of degeneration and for arbitrary nonparabolicities of bands, and the only approximations are performed in the following places: it is assumed that perturbation is quasi-stationary; a classic approximation for the Debye screening length is used; and Umklapp processes are neglected.

All of the quoted simplification introduce an error below 10 % into the calculated value of the Auger lifetime. Gerhardts et al. used the above expression to calculate Auger lifetime and obtained a satisfactory agreement with experiment [40].

There are various approximations in literature for intrinsic Auger lifetime for the CCCH process for the case of parabolic bands and nondegenerate semiconductor. The most often met is

$$\begin{aligned}\tau_{A1}^{i} =&(4\pi\varepsilon_0)^2\frac{\sqrt{2\pi}\hbar^3}{16m_0q^4}\varepsilon_s^2\left(\frac{m_0}{m_n^*}\right)\sqrt{1+\frac{m_n^*}{m_p^*}\left(1+2\frac{m_n^*}{m_p^*}\right)\left(\frac{E_g}{k_0T}\right)^{3/2}}\\ &\times\exp\left[\left(1+2\frac{m_n^*}{m_p^*}\right)\left(1+\frac{m_n^*}{m_p^*}\right)^{-1}\frac{E_g}{k_bT}\right]|F_1F_2|^{-2}.\end{aligned} \tag{1.51}$$

Here ε_S denotes the static relative dielectric permittivity of semiconductor, while ε_0 is the absolute dielectric permittivity of vacuum.

The coefficient of (1.51) is $(4\pi\varepsilon_0)^2\sqrt{2\pi}/(16m_0q^4) = 3.808 \times 10^{-18}$ (in CGS). Twice larger values for this expressions are sometimes met in literature (e.g., [19]) depending on the manner in which Auger lifetime was defined.

Thus there is a threshold energy for Auger processes, contained in the exponential term of the above expression (1.51). This energy is somewhat larger than the bandgap

$$E_T = \frac{1+2m_n^*\Big/m_p^*}{1+m_n^*\Big/m_p^*}E_g. \tag{1.52}$$

This is the minimum energy required for the electron $2'$ to start impact ionization.

The ration of intrinsic Auger lifetimes for Auger 1 and Auger 7 processes is approximately [41]

$$\gamma = \frac{\tau_{A7}^{i}}{\tau_{A1}^{i}} = 2\frac{m_n^*(E_T)}{m_0}\frac{1 - 5/(4E_T/k_bT)}{1 - 3/(2E_T/k_bT)}. \tag{1.53}$$

This approximation is valid for nondegenerate p-type HgCdTe $5 \times 10^{14} \le p \le 5 \times 10^{17}$ cm^{-3}, composition $0.16 \le x \le 0.3$, temperature $50\text{ K} < T < 300\text{ K}$ and small perturbations.

The total lifetime due to both CCCH and CHHL processes is

$$\tau_A = \frac{\tau_{A1}\tau_{A7}}{\tau_{A1} + \tau_{A7}}. \tag{1.54}$$

As far as solutions for a fully degenerate case are concerned, a satisfactory general analytical expression like the one for nondegenerate case (1.51) still does not exist. There are some approximations based on the assumption that Auger g-r rate decreases with an increase of carrier concentration, i.e., with a shift of the Fermi level deeper into the conduction band. White [50] empirically modeled this phenomenon by a multiplicative factor $(1 + \alpha_w n)$ introduced into the lifetime expression, where he used the value 1.9×10^{-17} cm^{-3} for the coefficient α_w.

Djurić et al. [51] used the fact that in degenerate semiconductors all states in the conduction band below $E_g + E_{\text{fn}} - 4k_bT$ (where E_{fn} is quasi-Fermi level for electrons) are occupied. Regarding interband transitions, such semiconductor behaves similar to an effective material with a bandgap widened for an additive factor of $E_{\text{fn}} - 4k_bT$. Thus, in their approximation the threshold energy of Auger processes becomes

$$\overset{\circ}{E_T} = \frac{1 + 2m_n/m_p}{1 + m_n/m_p}\left(E_g + E_{\text{fn}} - 4k_bT\right), \tag{1.55}$$

while intrinsic Auger lifetime in degenerate material is

$$\tau_{A1}^{i\,\circ} = \tau_{A1}^{i}\left(\frac{E_g + E_{\text{fn}} - 4k_bT}{E_g}\right)^{3/2}\exp\left(\frac{1 + 2m_n/m_p}{1 + m_n/m_p}\frac{E_{\text{fn}} - 4k_bT}{k_bT}\right). \tag{1.56}$$

Instead of the factor $E_g + E_{\text{fn}} - 4k_bT$ an alternative is to use the Burstein-Moss's effective bandgap. In the case of $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ it can be approximated by the following empirical expression [24]

$$E_{g\text{eff}}[\text{eV}] = E_g + E_{\text{MB}} = E_g + \frac{1 - x}{2}\left(\frac{n[\text{cm}^{-3}]}{7.85 \times 10^{18}e^{0.002T}}\right)^{0.63x+0.00038T+0.294}. \tag{1.57}$$

At very high concentrations, free carriers interact electrostatically with a consequence of bandgap narrowing and deforming. In that case, it is necessary to subtract from the above value for E_{geff} the terms taking into account this interaction.

Bandgap narrowing due to Coulomb interaction electron–electron may be written as [52]

$$\Delta E_{e-e}[\mathrm{eV}] = \frac{q}{2\pi\varepsilon_0\varepsilon_r}\sqrt[3]{3n/\pi}. \tag{1.58}$$

while Coulomb interaction between free electrons and impurities is [53]

$$\Delta E_{e-t}[\mathrm{eV}] = \frac{\pi^{4/3}\hbar^2}{3^{1/3}m_n^*}\left[1-2\left(\frac{4E_{\mathrm{MB}}m_n^*}{(3\pi^2)^{2/3}\hbar^{2/3}}-1\right)\right]. \tag{1.59}$$

An approximation for Auger lifetime in degenerate narrow-bandgap semiconductor was proposed by Beattie and White [54]. It was obtained by direct calculation according to the complete quantum mechanical model of Auger recombination, but with introducing simplifications. Since the valence band in the materials of interest is much wider than the conductive band (more than an order of magnitude), the authors introduced the approximation that it is completely flat. This enables decoupling of energy and momentum and significant simplification of integrals, with a result that the expressions for g-r rates and lifetime are only scaled by a single coefficient. The temperature dependence of this coefficient is determined by comparing it with accurate calculation. This approximation gives an analytical expression, relatively complex but highly accurate, with two fitting parameters to be determined using accurate model.

For each temperature and bandgap there is a threshold electron concentration for which the lengthening of Auger lifetime in comparison to nondegenerate case becomes marked. Figure 1.5 shows this concentration versus temperature, calculated here for $Hg_{1-x}Cd_xTe$ with E_g = 0.1 eV (i.e., for various cadmium molar fractions x).

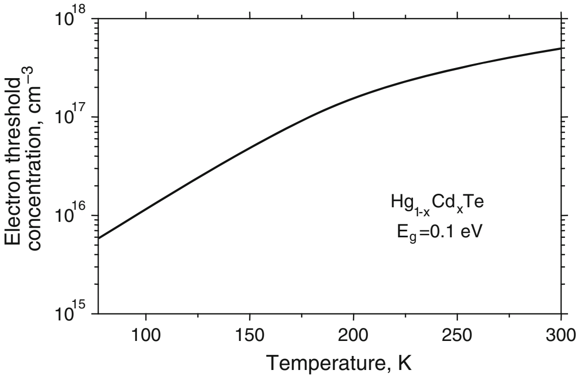


Fig. 1.5 Threshold electron concentration for Auger 1 process versus temperature in mercury cadmium telluride with a bandgap of 0.1 eV

Fig. 1.6 Temperature dependence of total Auger lifetime for $Hg_{1-x}Cd_xTe$ with a cadmium molar fraction $x = 0.195$ for different electron concentrations

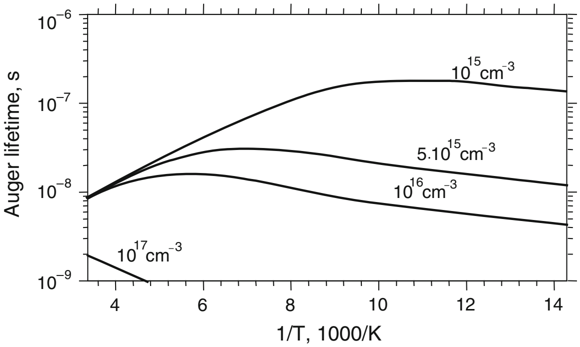


Figure 1.6 shows total Auger lifetime versus reciprocal temperature for various electron concentrations. The dependence was calculated according to (1.43), (1.45), and (1.48–1.51). Material was again mercury cadmium telluride, with $x = 0.195$.

1.4.3 Shockley-Read Processes

Shockley-Read (SR) [55] or Shockley-Read-Hall [56] generation-recombination processes proceed via imperfections ("traps"), i.e., centers of capture in semiconductor crystal lattice. Acceptor levels capture electrons and donor levels holes, or they emit them at rates dependent on the nature and the concentration of traps, as well as on the occupancy of energy levels. At that, within the bandgap there may be one or more impurity levels. SR mechanisms are more marked in technologically lower-quality material (with a larger concentration of defects and impurities), i.e., this mechanism is not fundamental.

Figure 1.7 shows the basic types of single-level Shockley-Read transitions.

An energy level of a trap with an energy E_t between E_c and E_v (index "t" means "trap") may capture an electron from the conduction band, emit an electron into the conduction band (thermal emission), capture a hole or capture an electron from the valence band. In the course of this process, electron energy may be converted into a phonon or a light quantum, depending on the nature of the center of capture. In calculation of this process we assume that the time between the transition of a

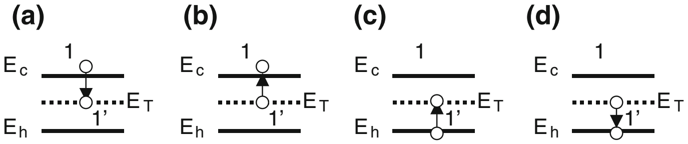


Fig. 1.7 Schematic presentation of single-level Shockley-Read transitions. **a** electron capture by trap; **b** electron emission by trap; **c** hole capture by trap; **d** hole emission by trap

carrier into the final state and its relaxation may be neglected in comparison of the duration of the recombination process itself.

Let us denote the Fermi function of probability for a capture center to be occupied by an electron by f_t, and the probability of the occupancy of a state in the conduction band by f_c. The recombination rate of an electron from the conduction band (in the energy interval dE around the value E) on traps in a unit volume of material is proportional to $N(E)\mathrm{d}E$ (concentration of electrons with energies within the interval under consideration) and N_t (concentration of vacant traps). Let $C_{\mathrm{SR}} = \langle v_n \sigma_n \rangle$ be the average probability per unit time for an electron from the interval dE to be captured at an empty center. Here $v_n = [8k_bT/\pi m^*]^{1/2}$ is thermal velocity of electron and σ_n is experimentally determined effective cross section of the electron capture at the center. The parameters σ_p and v_p are defined for holes in an equivalent manner. The rate of recombination at traps is

$$R_{\mathrm{SR}} = (1 - f_t)N_t \int_{E_C}^{\infty} C_{\mathrm{SR}} N(E) f_c \mathrm{d}E = \alpha_n N_t (1 - f_t) n. \tag{1.60}$$

Here α_n denotes the coefficient of capture of an electron at a center

$$\alpha_n = \int_{E'_C}^{\infty} C_{\mathrm{SR}} N(E) f_c \mathrm{d}E \Bigg/ \int_{E'_C}^{\infty} N(E) f_c \mathrm{d}E. \tag{1.61}$$

The probability of electron emission from a trap into an energy interval dE within the conduction band is proportional to the concentration of filled traps and may be represented by an expression in a form of (1.60) as

$$G_{\mathrm{SR}} = \beta_n N_t f_t, \tag{1.62}$$

where β_n denotes the coefficient of electron emission from a trap. In thermal equilibrium $R_{\mathrm{SR}} = G_{\mathrm{SR}}$, and thus

$$\beta_n = \alpha_n \frac{1 - f_0}{f_0} n_0 = \alpha_n n_1 = \alpha_n N_C \exp\left[-\frac{E_C - E_t}{k_bT}\right]. \tag{1.63}$$

Here n_1 denotes the equilibrium electron concentration for the case when the trap level and Fermi level coincide. In an equivalent manner p_1 is introduced, the equilibrium hole concentration for the same case

$$n_1 = N_C \exp\left(\frac{E_t - E_C}{k_bT}\right), \quad p_1 = N_v \exp\left(\frac{E_v - E_t}{k_bT}\right). \tag{1.64}$$

The capture constants may be written as

$$\alpha_n = \sigma_n v_n, \quad \alpha_p = \sigma_p v_p. \tag{1.65}$$

The total recombination rate is equal to the difference of net generation and recombination rates

$$r_{\mathrm{SR}n} = \alpha_n N_t[n(1-f_t) - n_1 f_t], \quad r_{\mathrm{SR}p} = \alpha_p N_t[p f_t - p_1(1-f_t)]. \tag{1.66}$$

In the case of thermal equilibrium the above two expressions become equal. Thus, we calculate the function of trap occupancy as

$$f_t = \frac{\alpha_n n + \alpha_p p_1}{\alpha_n(n+n_1) + \alpha_p(p+p_1)}. \tag{1.67}$$

SR lifetimes corresponding to completely empty traps for electrons and completely filled for holes are

$$\tau_{n0} = \frac{1}{\alpha_n N_t} = \frac{1}{\alpha_n v_n N_t}, \; \tau_{p0} = \frac{1}{\alpha_p N_t} = \frac{1}{\alpha_p v_p N_t}. \tag{1.68}$$

The occupancy function expressed by τ_{n0} and τ_{p0} is

$$f_t = \frac{p_1 \tau_{n0} + n\tau_{p0}}{(n+n_1)\tau_{p0} + (p+p_1)\tau_{n0}}. \tag{1.69}$$

Shockley-Read net recombination rate is

$$R_{\mathrm{SR}} = \alpha_n N_t(1-f_t) = \frac{n(1-f_t)}{\tau_{n0}}. \tag{1.70}$$

Shockley-Read net generation rate is

$$G_{\mathrm{SR}} = \alpha_n N_t n_1 f_t = \frac{n_1 f_t}{\tau_{n0}}. \tag{1.71}$$

For the case when $\Delta n \neq \Delta p$, under the condition $\Delta n - \Delta p \ll \Delta p, \Delta n$, we introduce an additive correction to the occupancy function, δf_t

$$\Delta n = \Delta p - N_t \delta f_t. \tag{1.72}$$

Since in stationary state $R_p = R_n = R_{\mathrm{SR}}$ and $f_t = \frac{1}{1+n_1/n_0} = 1 - \frac{1}{1+p_1/p_0}$, we obtain

$$\delta f_t = \frac{\tau_{p0}(n_0+n_1)p_0\,\Delta n - \tau_{n0}(p_0+p_1)n_0\Delta p}{(n_0+n_1)(p_0+p_1)\left[\tau_{n0}(p_0+p_1)+\tau_{p0}(n_0+n_1)\right]}. \tag{1.73}$$

The connection between excess concentrations Δn and Δp is

$$\Delta p = \Delta n \frac{\tau_{n0}(p_0+p_1)+\tau_{p0}(n_0+n_1)+\frac{N_t\tau_{p0}}{1+p_1/p_0}}{\tau_{n0}(p_0+p_1)+\tau_{p0}(n_0+n_1)+\frac{N_t\tau_{n0}}{1+p_0/p_1}}. \tag{1.74}$$

According to the definition (1.17), Shockley-Read lifetimes of electrons and holes are

$$\tau_{\mathrm{SR}n} = \frac{\tau_{n0}(p_0+p_1)+\tau_{p0}(n_0+n_1)+\frac{N_t\tau_{n0}}{1+p_0/p_1}}{n_0+p_0+\frac{N_t\tau_{n0}}{(1+p_1/p_0)(1+p_0/p_1)}}, \tag{1.75}$$

$$\tau_{\mathrm{SR}p} = \frac{\tau_{n0}(p_0+p_1)+\tau_{p0}(n_0+n_1)+\frac{N_t\tau_{p0}}{1+p_1/p_0}}{n_0+p_0+\frac{N_t\tau_{n0}}{(1+p_1/p_0)(1+p_0/p_1)}}. \tag{1.76}$$

If there is a larger number of impurity levels, SR lifetime is first calculated for each of the levels separately, and the reciprocal value of total SR lifetime is obtain as a sum of all separate reciprocal SR values [57]

$$\frac{1}{\tau_{\mathrm{SR}}} = \sum_k \frac{1}{\tau_{\mathrm{SR}_k}}, \tag{1.77}$$

where we define the parameters N_t, E_t, as well as σ_i, v_i ($i = n, p$) separately for each level. In the general case of a large number of levels the expression to be used is [8]

$$\frac{1}{\tau_{\mathrm{SR}}} = N_t \int_k f_c(k)v_n(k)\sigma_n(k)\mathrm{d}k. \tag{1.78}$$

1.4.4 Optical Generation

The optical generation term is in the form [58]

$$G_{\mathrm{opt}} = \alpha\Phi e^{-\alpha y}, \tag{1.79}$$

or

$$G_{\mathrm{opt}} = \frac{\lambda}{\mathrm{hc}}\frac{1}{lw}\alpha P_0 e^{-\alpha y}, \tag{1.80}$$

where l and w are the width and the length of the active area, Φ is incident optical flux and P_0 is incident radiation power.

Literature quotes more complex expressions for this term, taking into account multiple reflections at the front and at the back surface, real spectral dependences of reflection coefficient and (possibly) finite surface recombination rates at detector surfaces. For instance, in the case described by (1.8) optical generation becomes

$$G_{\mathrm{opt}} = \frac{\lambda}{\mathrm{hc}}\frac{P_0}{lw}\alpha\frac{1-R_1}{1-R_1R_2e^{-2\alpha d}}\left[e^{-\alpha y}+R_2e^{-2\alpha d+\alpha y}\right] \tag{1.81}$$

1.4.5 Influence of Background Radiation

In practical situations, it is impossible to avoid the influence of the background radiation. It acts on a detector through the same mechanism and in the same manner as the signal radiation, but is typically present at all wavelengths starting from zero and ending at the cutoff wavelength. The background usually radiates as a blackbody, i.e., its radiation obeys Planck's law for a body heated to some temperature T_{BG} (BG stands for Background), which is in a general case different than the detector temperature (T_d). The detector "sees" the background within a given field of view θ.

We assume that at a given wavelength the background radiation generation term has a form identical to that of the optical generation term. If we replace P_0 according to the Planck law and integrate the obtained expression over all allowed wavelength we obtain

$$g_{\mathrm{BG}} = \sin^2\left(\frac{\theta}{2}\right)\int\limits_{E_g/h}^{\infty}\frac{2\pi v^2}{c^2}\left[\frac{1}{e^{\mathrm{hv}/k_bT_{\mathrm{BG}}}-1}-\frac{1}{e^{\mathrm{hv}/k_bT_D}-1}\right]\alpha\frac{1-R_1}{1-R_1R_2e^{-2\alpha d}}\left[e^{-\alpha y}+R_2e-2\alpha d+\alpha y\right]\mathrm{d}v \tag{1.82}$$

1.4.6 Total G-R Rate of Narrow-Bandgap Semiconductors

Net recombination rate can be calculated as the sum of separate net recombination rates caused by independent mechanisms (denoted here by an index k), where each net recombination rate is equal to the difference between the generation and the recombination rate for the given mechanism

$$g_{\mathrm{tot}} = \mathrm{G}-\mathrm{R} = \sum_k g_k(n_0p_0-np). \tag{1.83}$$

If excess concentrations are small enough compared to the equilibrium values, (1.83) transforms into a sum of separate and independent generation-recombination rates defined by minority carrier lifetimes τ_k for each separate k-th mechanism [8]

$$g_{\mathrm{tot}} = -\sum_k \frac{\Delta n}{\tau_k}. \tag{1.84}$$

In the same approximation, the reciprocal value of the resulting lifetime is given as a sum of reciprocal values of each separate mechanism (a ratio between the generation-recombination rate and the excess concentration of carriers).

The total generation-recombination rate for narrow-bandgap direct semiconductors in the dark (i.e., without optical generation) is obtained as the sum of the rates for CCCH and CHHL Auger processes, radiative generation-recombination and single-level SR generation-recombination in the form

$$G = G_A + G_{\mathrm{rad}} + G_{\mathrm{SR}} = \left(\frac{n}{2\tau_{A1}^i} + \frac{p}{2\tau_{A7}^i}\right) + \frac{ni}{2\tau_{ri}} + \frac{(n_1 f_t)}{\tau_{n0}}, \tag{1.85}$$

$$R = R_A + R_{\mathrm{rad}} + R_{\mathrm{SR}} = \frac{np}{2n_i^2}\left(\frac{n}{\tau_{A1}^i} + \frac{p}{\tau_{A7}^i}\right) + \frac{np}{2n_i\tau_{ri}} + \frac{n(1-f_t)}{\tau_{n0}}. \tag{1.86}$$

If external IR flux is present, we add the optical generation term (1.79) or (1.81) into (1.85). If we take into account background generation as well, we add (1.82).

1.5 Noise in Semiconductor IR Detectors

In this section, we give a brief phenomenological description of noise in photonic detectors. Only the most basic practical details are included necessary for us to consider the possibilities to decrease it. An attempt has been made to give a theory valid simultaneously for photoconductors and photovoltaics.

We define intrinsic noise of an electron device as stochastic fluctuations of its electrical signal caused by internal processes in the device. The noise value in any moment will not be in correlation with its value in any other moment. A momentary value of noise cannot be measured, but only a sample in a given time interval determined by Δf. No electron device is completely noise-free. There are two fundamental reasons for noise appearance in photon detectors with an electric current flow: fluctuations of the number of carriers and fluctuations of the carrier velocity.

The main (but not the only) reason for the appearance of carrier number fluctuations are generation-recombination processes, i.e., the same processes that enable the very operation of the device and define both the dark current of the device and its optically generated signal.

Fluctuations of carrier velocity are a result of the Brownian nature of carrier motion within semiconductor material. They are a consequence of the fact that semiconductor material is at a temperature different than zero.

1.5.1 Thermal Noise

Let us first shortly consider velocity fluctuation noise. This is the only noise mechanism that exists without electric bias. It is variably denoted as Johnson noise, thermal noise, Nyquist noise, Johnson-Nyquist noise, “resistance” noise [59, 60]. It is a consequence of stochastic motion of charge carriers within material with finite resistance, and it represents a mechanism to maintain thermal equilibrium in semiconductor [61]. In a general case the spectral density of thermal noise voltage is $S_v(\omega) = 4R(\mathrm{hv}/2 + \mathrm{hv}/\left(e^{\mathrm{hv}/kT} - 1\right) = 2R\mathrm{hv}\,\mathrm{ch}(\mathrm{hv}/k_bT)$, which in the case $h\nu \ll k_bT$ becomes $S_v = 4k_bTR$, i.e., in this case the squared current due to this noise mechanism is

$$I_J^2 = 4\left(\frac{k_bT}{R}\right)\Delta f, \tag{1.87}$$

where R is semiconductor resistance. Thus, when $h\nu \ll k_bT$ thermal noise is “white”, i.e., its spectral dependence is flat over all frequencies.

1.5.2 Generation-Recombination Noise

Noise due to carrier number fluctuation is connected with detector bias and is denoted as shot noise or Schottky noise [5]. It is also denoted as the “excess” noise, but this expression is also sometimes used for $1/f$ noise [62]. It is a consequence of carriers passing through energy barriers, i.e., it appears as a result of the statistic nature of interband transitions and transitions band-impurity level, and in the final instance it is a consequence of the discrete nature of carriers [63]. When carrier number fluctuations are caused by g-r processes, this noise is also denoted as generation-recombination (g-r) noise.

For our present consideration g-r noise is the most important noise mechanism. The g-r noise spectrum is flat (white) up until the cutoff frequency, approximately given as the reciprocal value of free carriers lifetime. For the component of this noise not connected with illumination in a case of an ideal photoconductive detector one may write [7]

$$I_n^2 = 4q^2(|G| + |R|)\Gamma^2\Delta f, \tag{1.88}$$

while the shot noise for a photodiode is [7]

$$I_n^2 = 2q\left(\frac{I_d}{\beta}e^{qU/\beta k_bT} + 1\right)\Gamma^2\Delta f, \tag{1.89}$$

where $\beta = 1$ in the case of ideal p-n and Schottky diodes. U denotes bias voltage.

The above means that in both cases noise current is proportional to the total number of generation-recombination acts.

1.5.3 *1/f Noise*

The above consideration does not encompass the very important 1/*f* noise (flicker noise, current noise, modulation noise). It is marked at lower frequency ("pink noise") and becomes negligible in comparison to g-r noise at a "knee" frequency, a value that may be anywhere between several Hz and 100 MHz, depending on a particular device. In a general case flicker noise is given by the general empirical relation [64]

$$i_{1/f}^2 = K\frac{\langle i\rangle^{\alpha_{1/f}}}{f^{\beta_{1/f}}}\Delta f, \tag{1.90}$$

where K, $\alpha_{1/f}$ and $\beta_{1/f}$ are numerical constants, while i is current through detector. Typically $\alpha_{1/f} = 2$, while $\beta_{1/f} \approx 1$ (actually this value may be in the range 0.8–1.5).

A generally accepted theory of 1/*f* noise does not exist, although a large body of papers was published on this topic. Most of them start with the assumption that this type of noise is a consequence of stochastic fluctuations of either the density of free carriers or their mobility [61]. This noise can be, thus, connected with the effective scattering cross section in material and is often considered a consequence of the existence of potential barriers on the surface or in the bulk of semiconductor [8]. 1/*f* noise on the surface is caused by the transitions of carriers connected with "slow" surface states and with detector electrical contacts [64]. Often (but not always) this noise may be minimized by convenient technological procedures for surface treatment and Ohmic contact fabrication. Thus, it may be regarded more a technological than a fundamental problem.

1.5.4 *Total Detector Noise*

In this consideration we neglect all other noise mechanisms, e.g., noise caused by nonuniformity of impurity distribution within detector ("pattern" noise), all avalanche-related processes, burst ("popcorn") noise, etc. [65, 66].

Since all of the considered noise mechanisms are independent, the squared values of all of them are summed, giving the square of total detector noise.

Since photonic detectors most often operate at mid- and high frequencies, the most prominent noise component is g-r noise. Long [7] neglected 1/*f* component to give an expression for noise current valid both for photovoltaic (PV) and photoconductive (PC) detectors

$$i_n^2 = uq\left[q\eta\Phi_b A\Gamma^2 + I_d(V) + \frac{k_b T}{qR_d}\right]\Delta f, \tag{1.91}$$

where the factor u is 2 for PV and 4 for PC, photoelectric gain Γ is usually (but not always) equal to one for PV devices, while for PC it is almost always higher. Φ_b is background radiation flux, I_d is the dark current. In the case of PC detectors if sweepout effects are taken into account the first two bracketed terms should be multiplied by the sweepout factor, whose value is between 1/2 and 1.

Since Johnson noise is much lower than g-r noise in the operating frequency range of photonic detectors, from the point of view of noise optimization it is most important to minimize generation-recombination processes.

We consider further the determination of local values of generation-recombination noise for the case when carrier concentration within detector is position-dependent, i.e., when g-r rate and photoelectric gain are spatially inhomogeneous. We use such spatial distribution to determine total noise current (g-r plus thermal) through the whole detector. For the sake of simplicity, we assume that the gradient exists only along one direction, parallel to the y-axis. We consider a photoconductive device.

We divide the detector structure into a large number of layers sufficiently thin to permit us to regard g-r rate and photoelectric gain as constant. To one such layers with a thickness dy and located at a position y we apply the standard expression for the g-r noise of photoconductors (1.88). For a photovoltaic detector (1.89) is used in the same manner. Thus, we obtain the increase of noise current due to the processes taking place in the layer itself

$$\mathrm{d}i_{g-r}^2 = 2lwq^2\Gamma^2(y)\Delta f[|G(y)| + |R(y)|]\mathrm{d}y. \tag{1.92}$$

In the above expression, we sum the absolute values of generation and recombination rate, because they define statistically independent g-r acts, whose total number determines the noise level.

In (1.92) $\Gamma(y)$ represents the position-dependent photoelectric gain. In a position y it will be proportional to the ratio of the total detector current change and the generation rate change in the considered thin layer, so that we may write [67]

$$\Gamma(y) = \frac{1}{qlwd}\frac{\mathrm{d}I}{\mathrm{d}g_\mathrm{tot}}, \tag{1.93}$$

where g_tot determines the total generation-recombination rate obtained as the difference between (1.85) and (1.86). The contribution of (1.93) to the change of the total photodetector current is calculated by adding the change of the g-r processes rate in the layer under consideration to the generation term in the continuity equation and calculating the detector current.

The total g-r noise current will be given as an integral of d$i_\mathrm{g-r}^2$ over the whole detector thickness

$$I_{g-r}^2 = \frac{2\Delta f}{lwd^2} \int_0^d \left[|G(y)| + |R(y)| \right] \left(\frac{\mathrm{d}I}{\mathrm{d}g_{\mathrm{tot}}} \right)^2 \mathrm{d}y. \tag{1.94}$$

Further, we calculate the total noise current as a sum of the squares of generation-recombination noise current obtained in the above manner and Johnson-Nyquist (thermal) current

$$I_{N\mathrm{tot}}^2 = \frac{2\Delta f}{lwd^2} \int_0^d \left[|G(y)| + |R(y)| \right] \left(\frac{\mathrm{d}I}{\mathrm{d}g_{\mathrm{tot}}} \right)^2 \mathrm{d}y + \frac{4k_b T \Delta f}{R_d}, \tag{1.95}$$

where R_d is dynamic resistance of the detector for the given bias, equal to dU/dI.

Figure 1.8 shows the well-known shape of spectral noise in a typical photonic detector. At lower frequencies $1/f$ dominates; however, it drops quickly with frequency until it reaches a plateau on mid and high frequencies, dominated by g-r noise. Circles are experimental data according to [68]. Above a cutoff frequency this noise also decreases, leaving only Johnson noise.

1.6 Optimizing D^*F^* Product

1.6.1 Conditions to Reach Maximum D^*f^* in Photonic Detectors

Based on the consideration in Sects. 1.3–1.5 we may summarize the conditions for performance improvement of photonic detectors as follows

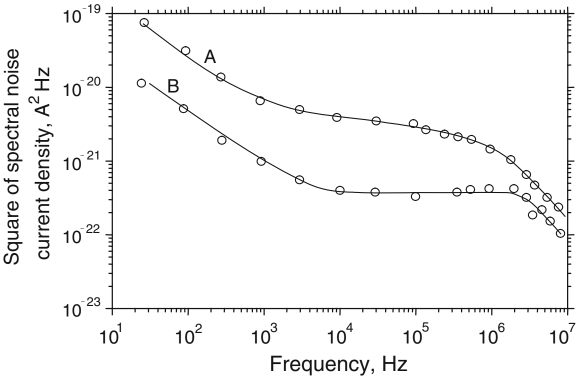


Fig. 1.8 Spectral noise density of a photovoltaic InSb detector with $1/f$, g-r and Johnson noise ranges. T = 120 K, surface recombination rate on the detector front side 100 cm/s, back side 10^4 cm/s. *Lines* theoretical, *circles* experimental. *Curve A* magnetic induction B = 0 T; *curve B* B = 0.18 T, Lorentz force directed toward the sample bulk

1. Operating wavelength of the detector should be as close to the cutoff wavelength ($\lambda_{co} = hc/E_g$) as possible. This requirement is easiest to meet in three-compound semiconductor materials with continually adjustable bandgap, e.g., mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) [8], mercury zinc telluride ($Hg_{1-x}Zn_xTe$) [69–71], lead tin telluride $Pb_{1-x}Sn_xTe$ [72, 73], and indium arsenide antimonide ($In_{1-x}As_xSb$) [74] which for $x = 0$ reduces to indium antimonide, InSb.
2. The part of the flux entering the detector should be as large as possible. This means that the detector structure should be optimized in order to maximize it, e.g., by decreasing the reflection coefficient of the incident surface and by maximizing the optical to electrical surface ratio.
3. The absorption coefficient of semiconductor material in the spectral range of interest should be as large as possible. For a given operating temperature this is a matter of a particular semiconductor and of the level of its doping.
4. The optical path of the useful LWIR or MWIR signal through the active area should be as long as possible. This is equivalent to the requirement to maximize quantum efficiency.
5. The detector thickness should be as small as possible. This requirement is opposed to the previous one. Thin detector structures enable a shorter response time (because of the shorter transit time) and lower noise levels (because detector volume is decreased).
6. Photoelectric gain should be maximized, but without compromising other detector characteristics like response time or noise level.
7. Detector noise must be minimized.

We will focus our attention to the last item, noise decrease in photodetectors. We start from (1.95), assuming that the given parameters are temperature (must be as near to the room temperature as possible) and photoelectric gain (must furnish maximum sensitivity and basically is given by the chosen detection mechanism). We conclude that the following two conditions are to be met

1. The dynamic resistance of detector in the operating point should be maximized. This is connected with the second term in (1.95). Since in principle R_d is inversely proportional to carrier concentration, the number of free electrons and holes in the active region of detector without illumination should be minimized. Usually the influence of g-r noise is much larger than the thermal part at usual operating frequencies (see the typical noise curve in Fig. 1.8) so that in practical situations the second term in (1.6) may be neglected.
2. The sum of absolute values of generation and recombination rate must be minimized.

If we look at the definitions of generation and recombination rates (1.85) and (1.86) we conclude that we may try to minimize Shockley-Read, Auger and radiative component of generation and recombination. It was already mentioned in Sect. 1.4.3 that SR processes are not fundamental and that a convenient fabrication technology could reduce them to a low enough level. On the other hand, Auger processes (and especially recombination) are strongly dependent on carrier concentration—see (1.38)—so that they rapidly rise with temperature and soon become dominant. This practically means that the main problem when minimizing the influence of g-r processes is how to suppress Auger processes. On the other hand, because of the strong dependence of the Auger recombination and generation rates on carrier concentration, the most rational approach is to decrease the number of electrons and holes in the active region. The conventional method is to cryogenically decrease the operating temperature, but it has been already said that this work is dedicated to the optimization of IR detectors in order to reach the BLIP operation without cooling.

If Auger processes are suppressed, the prevailing g-r mechanism remains the radiative processes. According to e.g., [3, 4] these processes exceed Johnson noise and remain the dominant factor determining the BLIP limit of device performance.

All of the requirements mentioned above may be conditionally classified into the technological ones (those regarding material composition and purity, number of defects in semiconductor material, geometry and surface treatment of detector, type and quality of contacts, etc.) and physical (connected in one hand to the increase of flux and optical path and generally to the light absorption in the detector, and in other to the decrease of carrier concentration in the active region with a goal to suppress noise.) Although the “technological” and the “physical” methods are similarly complex and practically of the same importance, this text is limited to the latter. This means that we assume that the detectors are fully defined from the technological point of view at a level optimal with regard to the state-of-the-art processes and procedures.

In our further considerations we start from the assumption that the geometrical and physical parameters of the photodetector are already defined. Geometrical parameters include the area and the thickness of the active region, i.e., its total volume. Physical parameters are the detector material properties. These include optical properties (complex refractive index, determined by the material type and composition), and electrical ones (donor and acceptor concentration, i.e., the concentrations of majority and minority carriers at a given temperature).

Our goal is to determine the possible approaches to optimize photon absorption and carrier transport in the active region of a semiconductor photonic infrared detector, and thus maximize specific detectivity. This should enable the elevation of the detector operating temperature and its operation as close to the room temperature as possible, all for a given spectral range and for a predefined set of technological parameters.

1.6.2 Generalized Photonic IR Detector

Based on the previous consideration, one can divide physical methods for photodetector performance improvement into two general groups. One of them are optical or equilibrium methods, and the other are nonequilibrium methods.

We denote as optical or equilibrium methods all various procedures that increase the number of photons of the useful signal within the detector active area. This means that they improve the performance of photonic MWIR or LWIR detectors without causing a nonequilibrium between charge carriers and the semiconductor crystal lattice.

The analysis of the expressions for specific detectivity and quantum efficiency shows that there are two basic approaches to this optimization. One of them is to increase the optical path of the incident radiation through the detector, corresponding to an effective increase of the active region thickness, but keeping simultaneously the real detector thickness as small as possible to minimize detector volume and thus noise. The second one is to increase the IR radiation flux entering the structure, which may be done by minimizing reflection and scattering on the incident surface and by some kind of incident flux concentration or focusing. Both of these belong to photon management or light management techniques.

Nonequilibrium methods are those that change the carrier distribution by elevating the effective temperature of carriers over the crystal lattice temperature. Thus, obtained nonequilibrium causes spatial redistribution of charge carriers, and can be thus utilized to decrease the carrier concentration in the desired part (the active area) of a photonic detector. This decreases carrier concentration-dependent g-r process rates, causes thermally-induced noise drop and as a result produce effects similar to those of cryogenic cooling.

The equilibrium methods can be used for photodetectors generally, while nonequilibrium methods are convenient for different types of semiconductor photonic detectors. Mutual for both equilibrium and nonequilibrium methods is that they can be implemented utilizing different micro and nanofabrication methods.

Figure 1.9 shows a general infrared photodetector, which utilizes all of the above enhancement methods. The general IR detector includes

(a) Optical concentrator—any structure serving to increase the portion of the incident optical flux entering the detector. This may include various light focusing units like immersion lenses, but also antireflection structures.

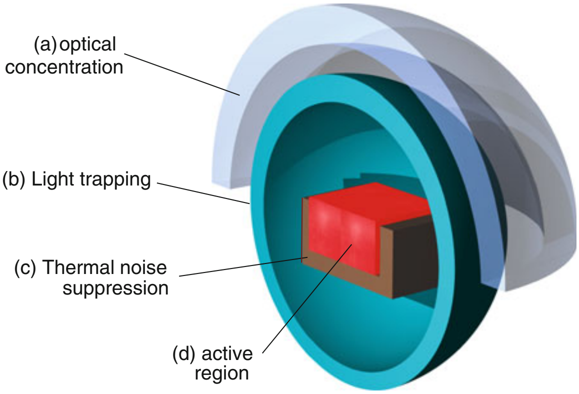


Fig. 1.9 Schematic presentation of a generalized infrared detector with the four main units

(b) Light trapping—structures increasing the optical path through the detector active area. At the same time, owing to the effect of reabsorption (photon recycling) light trapping increases the radiative lifetime of carriers within the device.
(c) Thermal noise suppression structure, intended to minimize Auger g-r processes, i.e., to decrease thermally induced fluctuations. In a general case it may include external cooling by liquid nitrogen Dewar or by some kind of active cooler. In the case of interests for this book, there is no cooling, and the role of cryogenic system is replaced by nonequilibrium structures operating at elevated temperature.
(d) Active detection zone—the part of the detector where conversion from the optical signal to another kind of signal is done, typically to a proportional current or voltage change. In the case of an intrinsic photonic IR detector it is usually the zone of interband carrier transitions.

The presented four main parts of a detector may be or may not be separate units. Actually, both their structures and their functions may overlap. For instance, the blocks (a) and (b) indirectly decrease noise by decreasing the active area volume. The block (b) also directly decreases noise since it increases reabsorption.

A real detection scheme may include all of the above or only some of them. Actually, only the (d) part—active detection zone—must exist in all photodetectors, while all of the others are optional and their function is to improve overall performance by increasing the values of different factors in the specific detectivity-bandwidth product.

The rest of the book is dedicated to the above quoted methods of photodetector enhancement. It is divided into two parts. One of them analyzes different methods to optimize the photodetector input, i.e., to perform its optical improvement, while the other part is denoted to nonequilibrium methods of noise suppression.

Chapter 2
Photon Management

2.1 Fundamentals of Photon Management in Photodetectors

2.1.1 Introduction

The optical methods of optimization have a goal to maximize the quantum efficiency of a photodetector for a given geometry (thickness and active area) and a given spatial and frequency dependence of the absorption coefficient. They are described as photon management or light management methods.

Various photon management methods considered in this chapter are mostly generally applicable, not only for infrared semiconductor devices. We consider a general case of a photodetector as a device that converts optical energy into another form of energy. Most often this energy is electrical signal, although other forms may be used like thermal [6], motion (like in some microcantilever-based detectors) [75], optical signal at another frequency (up- or down-converted) [76, 77], etc. A notable example of devices to which a majority of the methods described here can be applied are thin-film solar cells [78–81].

There are different methods of light management in a photodetector that can be divided into four groups, Fig. 2.1: optical concentration, use of antireflection structures, optical path increase, and light localization.

Optical concentration methods actually collect incident radiation from a larger area (denoted as "optical" area) and concentrate it (focus) to the smaller active area of photodetector (the "electrical" area). The concentration efficiency can be then defined as the ratio between the optical and the electrical area, minus absorption and scattering losses. This method of light management is typically done by utilizing structures which are not themselves a part of the detector, but can be integrated with it. The simplest case of a concentrator would be an immersion lens, but there are a number of different other structures to serve the same purpose. Roughly, one could

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2_2

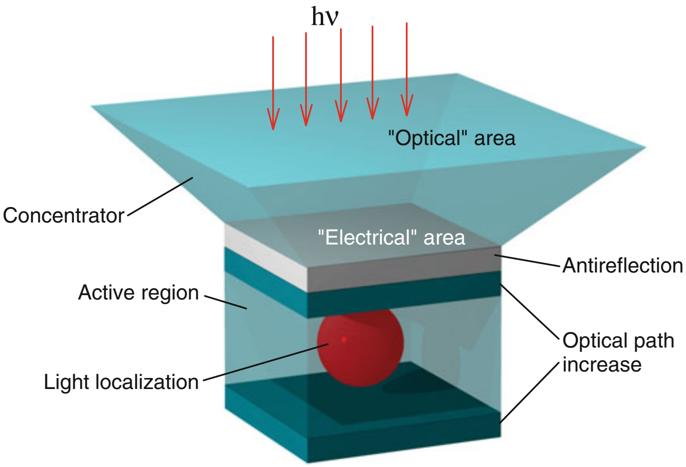


Fig. 2.1 Methods of photon management: use of optical concentrator, antireflection structure, structures for optical path increase (cavity enhancement) and light localization structures

divide them into refractive, reflective, and diffractive elements, although hybrid solutions are also possible.

After arriving to the detector surface, the incident beam should enter the photodetector area with losses as low as possible. Semiconductor materials used for photodetectors have large values of refractive index and thus large values of reflection coefficient at the device surface. This reflection is minimized using antireflection structures. Basically, these structures serve as impedance matching media between the detector environment (most often, but not always free space) and the active region.

Once useful radiation enters the detector-active region, the goal is to keep it there as long as possible and thus obtain maximum interaction between it and the active material. Probably the simplest way to do that is to place a mirror at the back side of the detector, thus doubling the optical path through the active region. However, the methods of optical path increase can be much more sophisticated and may include different cavities with reflective walls, as well as surface structuring. These methods belong to the optical trapping approaches.

There is another method of optical trapping, enabled by the advent of nanophotonics: the subwavelength optical localization utilizing plasmonic nanocomposites. Some metal-dielectric structures ensure the possibility of light localization on a level much smaller than the operating wavelength. To this purpose they utilize the propagating or localized surface plasmons polaritons.

This chapter investigates in some detail all of the mentioned methods of photodetector enhancement.

2.1.2 Fundamental Limits of Photon Management

Basically, all photon management techniques are intended to improve absorption of light in the detector, thus ensuring a higher degree of the conversion of incident optical radiation into a useful signal. Obviously, the efficiency of any conversion is limited by basic physical laws. A question is posed what are the fundamental limits of photodetector enhancement through light management.

A detector system is presented in Fig. 2.2. A background optical flux is incident to the active area of a photodetector with a thickness d. Both in the case of solar cells and night vision photodetectors, the optical flux is blackbody radiation, described by the Planck's law. In a general case, the detector material may incorporate nanostructuring that could localize optical field and create hotspots with high density of states. A perfect mirror is placed at the rear side of the device—i.e., it is assumed that the incident light is unidirectional, while the internal radiation is bidirectional.

The detector surface is corrugated in order to increase the optical path through the detector. The corrugation may be random or ordered, but in both cases its basic purpose is to change the direction of light incident upon the active surface and to make use of total internal reflection to ensure repeated passing of the beams through the active region. Light can escape if the direction of the internal beam falls within the escape cone, for which according to Snell's law $\sin \theta_{cr} = 1/n$ (θ_{cr} is the critical angle of total reflection, n is the refractive index of the active region).

We first consider the case limited by geometrical optics, which has been established by Yablonovitch [82–85]. In the literature it is variably denoted as the conventional limit, the ergodic light trapping limit, the ray-optics limit and the Lambertian limit. It is assumed that the detector-active material can be described by an effective absorption coefficient α isotropic throughout the device and that the detector thickness is much larger than the operating wavelength in free space ($d \gg \lambda/2n$), so that one considers a bulk process. The absorbance within the photodetector for a single pass across the structure (absorption without enhancement) is

$$A(\omega) = 1 - \exp(-\alpha(\omega)d) \approx \alpha(w)d, \tag{2.1}$$

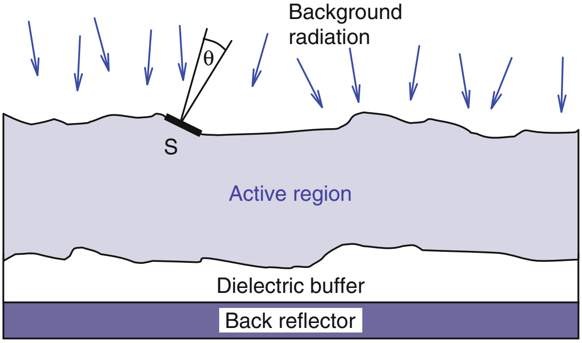


Fig. 2.2 The structure of the active region of a detector with corrugated surface and ideal backside reflector

i.e., the absorbance is equal to the optical thickness of a photodetector, which is defined as the $\alpha(\omega)d$ product.

Since a bulk case is considered, it is further assumed that interference/diffraction effects can be neglected and that the intensity of light within the detector medium is in equilibrium with external blackbody radiation. The density of states within the medium is proportional to n^2. The next assumptions are that the equipartition theorem is valid (the internal occupation of states is equal to the external one, the internal states are ergodic) and that the surface corrugation performs a full randomization of the incident signal over space. This is not always satisfied, but the assumption holds in a vast majority of cases [82]. A sufficient condition for randomization of light by multiple scattering corrugated surfaces is that these surfaces upon averaging behave as Lambertian. The internal distribution of the light within the medium is then isotropic.

According to the statistical ray optics approach [82], the relation between internal and external intensity of light is

$$I_{\text{int}}(\omega, x) = 2n^2(\omega, x) I_{\text{ext}}(\omega), \tag{2.2}$$

The same result is also obtained according to the principle of detailed balancing of the light [86] applied between the light incident to a small surface element of the detector-active area and escaping from that same element through the loss cone and by applying the brightness or radiance theorem (e.g., [87]) stating that the spectral radiance of light cannot be increased by passive optical devices (based on the principle of reversibility).

To determine the enhancement of absorption, one has to consider the loss of light due to various mechanisms. According to Yablonovitch [82, 83], there are three such mechanisms: the escape of light through the light cone, the losses due to imperfect reflection at the surfaces, and the absorption in bulk. The absorbance of a photon is the ratio of the rate at which absorption occurs and the sum of the absorption and the photon loss through the escape cone. For the volume absorption in the limiting case when $\alpha(\omega)d \ll 1$ and taking account the angle of the loss cone θ, this expression is

$$A(\omega) = \frac{\alpha(\omega)}{\alpha(\omega) + \frac{\sin^2\theta}{4n^2 d}}, \tag{2.3}$$

so that the absorption enhancement limit in the bulk case with internal randomization becomes $4n^2$. For $\theta = \pi/2$, this assumes the more often used simple form

$$A(\omega) = \frac{\alpha(\omega)}{\alpha(\omega) + \frac{1}{4n^2 d}}. \tag{2.4}$$

A more complex situation is encountered with the devices using plasmonic localization for absorption enhancement. In this case, many of the above assumptions introduced for ergodic limit are not valid. The crucial points are that the light distribution is not isotropic anymore (and actually the volumes with a strongly enhanced density of electromagnetic states may be deeply subwavelength) and the thickness of the detector is usually subwavelength.

A number of treatises are dedicated to the situations in which the ray optics limit is exceeded and optical modes are confined at subwavelength scale [88–90]. However, until now no generally valid solution has been given for the extension of the ray optics limit [91].

2.2 Nonimaging and Imaging Optical Concentrators

One of the obvious methods to increase the optical input to the detector is to collect optical signal from a larger area and to concentrate it to the smaller electrically active area. This can be done by some kind of optical concentrator, i.e., focusing optics integrated with the detector unit.

The usual way to concentrate light from a larger area to a smaller one is to utilize optical elements that can create an image of an object. However, besides such structures, there are also concentrators that belong to the wide class of nonimaging optics [92–97]. This is actually the branch of geometrical optics in which the main goal is not to form the best possible object image, but rather to gather as large portion of the incident radiation as possible and to concentrate it to the active area of the detector, regardless of the price to be paid from the point of image fidelity. A single photodetector element in infrared technique is not an imaging device and only the total intensity of the radiation impinging on its active surface is important, not its spatial distribution. A similar situation is met in the case of solar cells. Thus for such photodetectors one can apply both imaging and nonimaging concentrators.

A schematic presentation of an optical concentrator is given in Fig. 2.3. The shape of the system can be arbitrary, as well as its mechanism of focusing. The important factor of the system is its concentration ratio, defined as a ratio between the input (collector) area and the exit area. In an ideal concentration system, this ratio should be equal to the ratio of incident optical flux to the exit flux. In a real system, possible absorption losses (for instance in metal parts) will have to be deducted. Thus the concentration ratio of a light concentrator can be defined either as the geometric concentration ratio C, the ratio of the entry aperture area ($A_{\text{collector}}$) to the exit aperture area (A_{exit}), or as the irradiance gain C_i, the ratio of the irradiance on the collector (Ir_c) to that on the entry aperture (Ir_m). The two concentration terms are related by the fraction of total incident power entering the module that reaches the concentrator exit aperture.

Different strategies can be used to concentrate the incident optical power. There are three main groups of optical concentrators. One of them is based on refractive optics (conventional lenses). The second one is reflective concentrators (mirrors), while the third group is diffractive optical elements. Obviously, a system may simultaneously incorporate two or even all three of the mentioned structures.

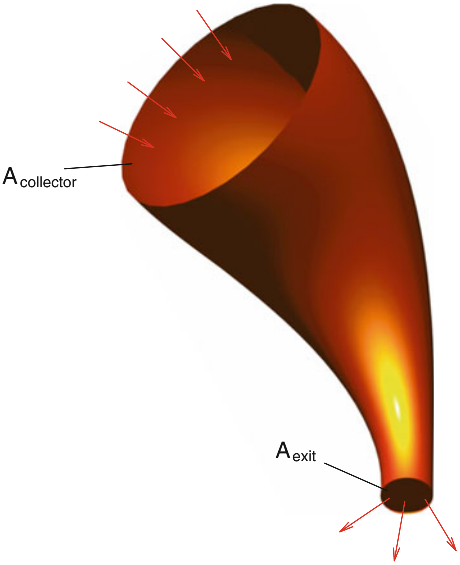


Fig. 2.3 Schematic presentation of an arbitrary light concentration system. The area at the entry of the system is $A_{\text{collector}}$, and at the output A_{exit}

Monolithic or hybrid integration may be used to integrate concentrators with the photodetector (optical immersion of detectors and focusing optics). The concentrators may be partially integrated with the detector, i.e., positioned immediately to the detector surface, but without physical integration. In that case, there is a gap filled with air or some transparent buffer material with a real part of its refractive index comparable to that of the detector-active area or smaller than it and at a distance of an order of the operating wavelength (pseudoimmersion). Finally, concentrators may be integrated with the detector housing, i.e., encapsulated together with the active element, but separated from the detector surface by a distance much larger than the operating wavelength (discrete lenses).

From the point of view of coupling efficiency (minimization of reflection losses), the best solution is monolithic or hybrid integration; however, in practical situations one can encounter all of the mentioned approaches. Two main types of focusing lenses may be used—either refractive lenses fabricated in material with high real part of refractive index and low absorption coefficient at IR wavelengths, or diffractive lenses. Any of them may be either discrete or arrayed. Reflective optical concentrators may be used, also reflective holographic optical elements or any of their combinations.

Besides the maximum increase of incident flux, the main requirements posed to the concentrators in infrared photodetection are the simplicity of fabrication technology, minimal losses, and good reproductivity during fabrication. The dimension tolerances for optics are of secondary importance and allowed aberrations are higher than those posed, e.g., to the microlenses used with optical fibers. Even in imaging optics for the infrared range the tolerances (especially those for MWIR and LWIR) are much less stringent than those for visible light, since larger wavelengths are used. The nonimaging nature of the concentrators further relaxes the tolerances.

2.3 Refractive Concentrators

2.3.1 Immersion Microlenses

Probably the most obvious way to collect illumination from a larger area and to concentrate it to the active area of a photodetector is to use a microlens. Such a lens should have dimensions sufficiently small to be mounted on a detector. Since detector diameters are typically below 1 mm and very often below 0.1 mm, this means that the “micro” in the name of the microlens only means “small,” since the dimensions of the lens itself will be of the order of millimeters.

In order to avoid reflective losses between the detector and the lens, typically some kind of refractive index-matching immersion will be used. The lenses for the photodetector improvement can be thus called immersion lenses. Obviously, the material of the lens has to be transparent in the wavelength range of interest and is typically used with some kind of antireflective coating.

There are a number of different geometries convenient for immersion microlenses in photodetection. Probably, the most well known and widely used form is hemisphere. The use of microsystem technologies allows the fabrication of various discrete or arrayed microlenses, with spherical surfaces (calottes, hemispheres and truncated spheres, full spheres), aspheric (ellipsoids, paraboloids, cylinders, cones), toroid, as well as various nonmonotonic surfaces consisting of two or more monotonous segments. Most of the microlenses convenient to increase the incident flux to the detector are plano-convex ones.

Figure 2.4 shows some types of refractive microlenses that can be fabricated utilizing the standard microfabrication procedures in materials convenient for the MWIR and LWIR ranges. Most of them are loosely based on the solutions for microlenses used in fiber optics to improve coupling between laser sources and fibers [98]. These immersion lenses were thus intended for the operation with coherent and monochromatic radiation, while most of the microlenses in the field of IR detector technology are intended for incoherent, mono- or polychromatic Lambertian sources and, of course, they operate in different atmospheric windows.

Among spherical microlenses utilized in optoelectronics literature quotes spherical calottes [99] (Fig. 2.4a), hemispheres [100, 101] (Fig. 2.4b) and truncated spheres [102] (Fig. 2.4e). These lenses may be defined as the refractive surfaces that can be represented as a part of a sphere with a radius r and with its center in the focus of the lens. The immersion lenses are usually formed by bringing into contact the planar surface of such a lens and the incident surface of the detector. Besides the above-mentioned lenses, ball-shaped microlenses are also used [103].

The following microlens shapes were also fabricated by microsystem technologies: elliptic calottes [104], hemiellipsoids [105], ellipsoids [103], hemicylinders [101, 105] (Fig. 2.4g), and cylinders [106] (Fig. 2.4h). These types of microlenses are especially convenient if the beam shape is elongated. Also among aspherics are conic microlenses with straight or curved generatrix [107], (Fig. 2.4i). In some optical systems, paraboloid or hyperboloid surfaces are sometimes used, usually

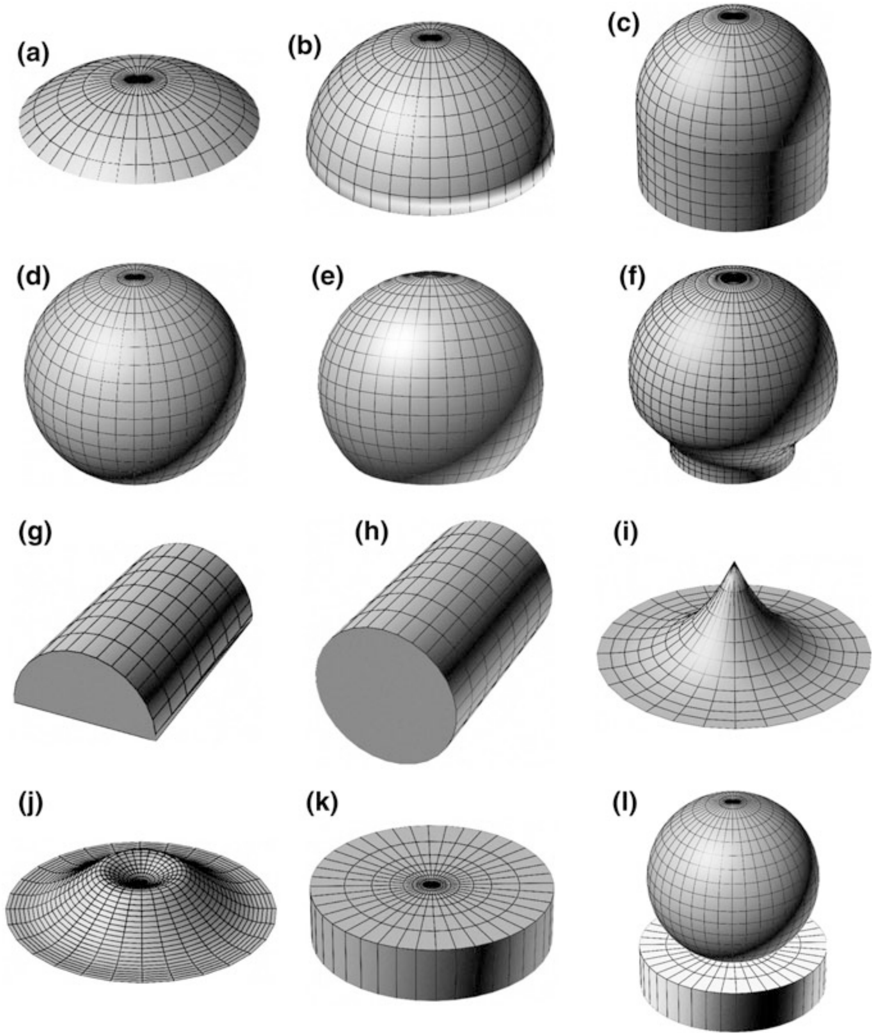


Fig. 2.4 Some spherical and aspheric discrete microlenses for optical concentrators in MWIR and LWIR range that may be fabricated by microsystem technologies. **a** calotte; **b** hemisphere; **c** hyperhemisphere; **d** ball lens; **e** truncated sphere; **f** "bulb"; **g** hemi-cylinder; **h** cylinder; **i** curvilinear cone; **j** concave concentrator; **k** gradient-index lens (GRIN); **l** complex two-element lens (sphere/GRIN). The grid corresponds to the homogeneity of refractive index, i.e., describes its gradient

fabricated by polishing. We should mention here optical concentrators with concave middle area [98] (Fig. 2.4j).

A very important type of optical concentrators, both in optical telecommunications and for MWIR and LWIR photodetectors are hyperhemispheric microlenses (combinations of a hemisphere and a cylinder with joined bases, i.e., the "rod lens")

[103] (Fig. 2.4c). Literature also mentions microlenses shaped as rounded pyramids, fabricated by micromachining [108].

An important class of microlenses are planar lenses with gradient refractive index (GRIN-microlenses, Gradient Index), Selfoc-lenses or Luneberg's lenses [103] (Fig. 2.4k; the grid shows the nonuniformity of refractive index, high in the middle, and decreasing toward the edges.) Within this context one could also mention the "electronic lenses" where the refractive index profile depends on impurity concentration and thus allows an external electrical control of the focus [109].

Confocal systems with two microlenses are often met in telecommunications [110], most often a combination ball lens-GRIN (Fig. 2.4l). Typical procedure is to focus one of the lenses and to align the other subsequently, so that the combination itself improves the allowed position tolerances several times. Also, for some applications (irregular beam shape) it is convenient to have more than one lens to allow a larger degree of freedom when adjusting beam collimation. Dependent on application, systems with three or more microlenses are also used [110].

2.3.2 MEMS Fabrication of Infrared Microlenses

As mentioned before, most of the structures for infrared detection can be fabricated by microsystem technologies (MST) and the microlenses for optical concentration are not an exception, regardless of the type of the optical element in question (refractive, reflective or DOE/HOE).

A point of special importance is the choice of proper material for MWIR and LWIR optical concentrators. A number of materials are customarily used to fabricate infrared optics. Besides these, optical concentrators may be fabricated directly in the photodetector material and indeed be monolithically integrated with it.

The usual requirements posed for infrared optics and regarding mechanical and water resistance, robustness and durability, insensitivity to temperature changes are significantly less stringent here, owing to the fact that microlenses are integrated within the detector housing. Some water-soluble materials are used for IR concentrator microlenses, because they are readily polished with water and cut by a wet string saw [1]. Among the materials used for infrared is germanium, since it ensures relatively simple machining. The most important requirements remain a maximized transmission in a given spectral range and the possibility to machine the microlens by some of the MST procedures.

A number of procedures are used to fabricate microlenses for telecommunications and some of them can be modified and adapted to fabricate optical concentrators for the MWIR and LWIR range. Some of these procedures already became standard.

Chronologically, the oldest approach to microlens fabrication is thermal melting. The action of heat is used to soften or melt convenient material (e.g., germanium in LWIR range) and form it into a desired shape. An alternative is to use material that polymerizes into a desired shape, or a convenient solver may be used that evaporates and leaves a solid object. Melting is the technology utilized for the least demanding applications, but at the same time it is probably the most often used one.

One of the oldest methods used with molding, but at the same time the least accurate, is cast molding (e.g., injection molding, casting, etc.) [111]. This method is used both for refractive and diffractive lenses. First, a high-accuracy "master" is fabricated. It is subsequently used to produce large series of microlenses. The method is especially convenient for work with plastic materials transparent in IR, like TPX (polymethylpentane, polymerized dimer of propylene.)

The procedures to fabricate masters are handled later in this section. Practically, all higher accuracy methods used to fabricate microlenses can be used to produce masters. In all of the quoted methods, it is important to choose polymers transparent in the IR range.

The methods for inexpensive replication and mass production of microlenses include

- Conventional injection molding (the most often used technique for low-end diffractive elements, for instance for the best-known diffractive optical elements, CD, DVD, and BluRay disks) and epoxy molding (the latter being more expensive because of longer curing cycles).
- Embossing: heat and pressure are applied to transfer surface relief from a mold into a plastic substrate, typically a thermoplastic foil.
- UV replication: probably the most convenient method for DOE mass production and also the most usual for higher quality and low-cost diffractive elements. The substrate is usually glass, and the mold is fabricated by electroforming. UV curable resin is potted between the mold and the glass and replicas are molded by photopolymerization. Alternatively, the mold may be used for e-beam evaporation of various inorganic materials to the substrate.

One of the molding methods often used to fabricate refractive lenses is to deposit a cube of lens material on the substrate using, e.g., some of the planar technology procedures, and then to treat it thermally until it melts. Surface tension turns it then into a hemisphere, calotte, or a flattened hemisphere [112]. An even better method is to use a similar procedure to deposit a "base" for the microlens with a desired diameter, and then to deposit lens material on it [113]. After melting, the lens material flows exactly to the edge, where surface tension prevents it to go farther (melt stop.) In this way, the lens surface is formed with an extreme accuracy.

One of the often used methods is to dip the lens support into the melt and remove it. In the course of this, a "drop" is formed at the end of the support. After cooling and hardening it is mechanically cut away. Finally, one of the used approaches is to drop the melt into cooling liquid. The shape and the size of the cooled drop define the lens.

The next important technology is photolithography. It is used both to fabricate lenses ready for use and masters for further replication, for any of refractive, reflective, or diffractive concentrators.

The most often used lithography procedures are electron-beam writing and optical lithography. Electron-beam writing is actually scanning of the photoresist-covered (spin-coated) surface by a focused electron beam, and thus exposing the resist. This technique enables the fabrication of continuous and curved profiles. It poses problems with positioning, requires computer control, and very high mechanical and vibration stability and must be performed in vacuum. Optical lithography uses light beam for writing and may be performed in the air. Some of its variants are direct laser writing in photoresist (laser beam lithography), holographic lithography (the use of interference pattern to write a DOE/HOE without a mask), and conventional photolithography (where mask can be also a hologram). The latter can use contact-printing (direct contact between the mask and the resist surface, furnishes the best results if performed in vacuum), proximity printing, and projection printing (conventional for volume production, but less accurate).

The next step in all the versions of lithography is removal of the photoresist, either the exposed part (positive photoresist) or unexposed (negative p.) and the final step is etching.

Two dry etching methods are often used: Reactive-Ion Etching (RIE) and ion milling. RIE is actually a simultaneous etching by chemical reaction and by particle bombing. It is anisotropic (vertical component is much more pronounced than the lateral). Ion milling uses a stream of inert gas to strike the surface and remove material by physical sputtering.

Wet etching includes chemical micromachining (etching in a chemically aggressive solution) and, depending on the properties on the substrate and the etching solution, it can be isotropic or anisotropic. Among the available techniques is laser-assisted chemical etching (convenient for very deep vertical holes). Blazing can be done by the melted resin technique and the mass transport method. Masters for simpler kinds of diffractive concentrators can be formed by holography or by mechanical ruling using, for example, a diamond cutter.

The fabrication of hemispheres and hemicylinders was reported in Cohen (1974), Bear (1980), Lee (1985). The microlenses here were fabricated in negative photoresist by point UV illumination to obtain a hemisphere, or by bar-like illumination for hemiellipsoides and hemicylinders. The excess resist is removed by solvent. A negative property of this method is that resist lenses are mechanically sensitive, stronger sources (e.g., industrial lasers) may easily damage them and they are applicable only for certain wavelength ranges.

This method is convenient for fabrication of GRIN lenses, where the required refractive index gradient is obtained by point diffusion [114]. Refractive microlenses with arbitrary geometry may be fabricated by photolithography so that first masks and etching are used to form a mesa or multiple-step structure directly in the substrate, and then additional etching is used (bromine-methanol for HgCdTe or InSb) to round it into a spherical or aspheric surface [115].

Another approach to form a lens directly in the substrate material is to form first a lens in photoresist deposited onto the substrate using some of the above procedures. After that reactive-ion etching (fast, but relatively small choice of materials) or ion milling (slower, more materials) is used to transfer lens into the substrate

[116]. The method is convenient for fabrication of optical elements in the wavelength ranges where resist is not transparent and generally offers a larger degree of freedom in choice of lens material. Finally, one of the options is to immerse the whole system into an electrolyte and to perform photochemical etching through a mask [117].

Most of the above photolithographic procedures can be utilized to fabricate diffractive lenses [118].

We mentioned holographic lithography as one of the types of photolithography used for IR concentrator fabrication [119]. Its main advantage is that the key steps of forming and exposing a mask are omitted. Resist (or some other medium, e.g., bichromated gelatin [120]) can be directly exposed by an interferometric pattern obtained by laser irradiation [107, 121]. A stable optical bench is necessary for that (or, alternatively, a pulse laser with high enough output power). The second alternative is to use computer-generated holography. In that case, a mask is necessary, but very complex systems can be fabricated [122]. Some authors mention the use of bichromated gelatin for the infrared optical elements [3].

Chemical micromachining is one of the methods of choice for fabrication of microlenses [123]. Silicon and germanium were used as lens materials. For spherical and aspheric geometries solutions for anisotropic etching are used, while nonmonotonous surfaces can be fabricated by combining masks and anisotropic etching. This method was used to fabricate various types of optical concentrators, among them cones and inflected surfaces [107].

Single microlenses with relatively larger dimensions can be efficiently produced mechanically, e.g., using abrasion or polishing. To fabricate both refractive and diffractive microlenses diamond turning is used [124]. One of the methods in this group is microgrinding [125].

One of the modifications of this approach is the use of a combination of mechanical and chemomechanical polishing, for instance, by simultaneous action of abrasive and bromine-methanol for lenses in Ge, CdTe, and InP [126].

2.4 Reflective Concentrators

Reflective nonimaging concentrators are not often used with infrared photodetectors, but are very well known and intensively applied in solar cell energy production [93, 123, 127–129]. Typically their structure is a tapering hollow with reflective walls, its wider aperture serving as the collecting surface, while multiple reflections from the walls collect beams toward the exit opening. The cavity may be empty (i.e., filled with air) or filled with some dielectric. The geometry may be two-dimensional (various kinds of trough) or three-dimensional, the latter being either a body of revolution or some other kind of tapering structure. The prototypical 3D structure is the so-called Winston cone, while various other tapering profiles may be used as well. They start from the simple ones like parabolic, hyperbolic, circular, conical shapes, extend to their combination, but may have different complex forms

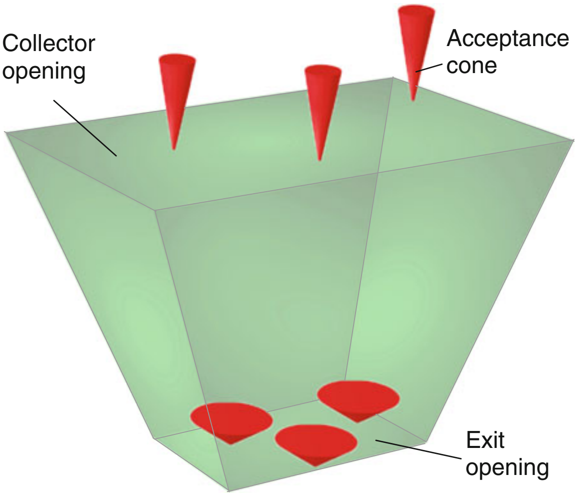


Fig. 2.5 Angular distribution of light showing larger acceptance angle cones at the exit opening and smaller at the collector opening

as determined by a particular application and the optimization procedure implemented to obtain maximum performance. The surface also may be multipart/segmented, with each segment having a different profile. Other solutions include Fresnel or echelon reflectors.

Figure 2.5 shows the connection between the spatial angle from which incident flux may arrive in order to be concentrated (acceptance cone) and the spatial angle under which radiation leaves the concentrator. Any radiation outside the acceptance cone will be rejected, i.e., the system will not bring it to the exit aperture. It can be seen that the acceptance cones at the input cover much smaller spatial angle than the light cones at the exit.

Regarding materials, the structures of microreflectors and their arrays are usually made of metal with good reflection coefficient. There are a wide choice of metals and alloys convenient for the fabrication of reflector-type concentrators.

The shape of a 3D reflector top and basis may be rounded, but also polygonal, the latter being more convenient if a high filling factor is desired when producing 2D arrays of reflectors. Figure 2.6 shows some possible shapes and the manners of their stacking into larger matrices. The rightmost example shows that the possible shapes may include combinations of polygonal entrance with round exit.

2.4.1 CPC (Winston Collector)

Compound parabolic concentrator (CPC, also called Winston collector or Winston cone) [95–97] is a reflective nonimaging light concentrator intended to collect incident light from a spatial angle larger than that of imaging concentrators and to

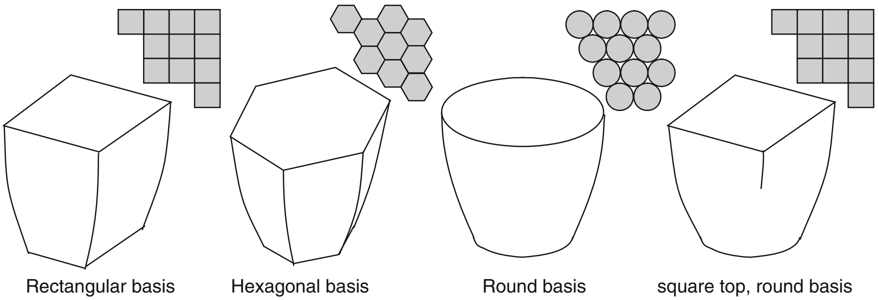


Fig. 2.6 Different entrance and exit geometries shapes for reflective light concentrators. *Gray matrices* in the right upper angle of each shape show the manner of concentrator packaging into arrays

direct it toward its exit opening. It has been the first discovered nonimaging reflective concentrator. Its 2D depiction is schematically shown in Fig. 2.7.

Two parabolas p_1, p_2 can be seen in the figure, with their respective focal points in F_1, F_2, and their axes being a_1, a_2. The parabolas are cut at the line of their focal points and then mutually rotated so that the focal point of one of them coincides with the curve of the other parabola and the very edge of the exit opening. A 3D version of this structure is obtained as a body of revolution whose sides are parabolic segments satisfying the same condition. Thus a Winston cone is an off-axis parabola of revolution.

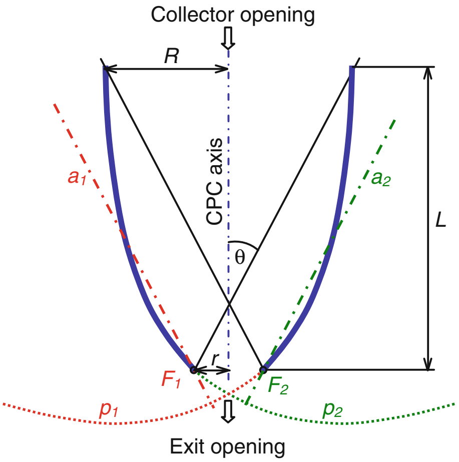


Fig. 2.7 Schematic presentation of CPC collector. R is the entry (collector opening) radius, r is the exit opening radius, a_1 is the axis of the parabola p_1, a_2 is the axis of the parabola p_2, F_1 and F_2 are focal points of the parabolas 1 and 2, respectively. Parabolas are rotated so that the focal point of one of the parabola coincides with the curve of the other parabola at the edge of the exit opening

A CPC concentrator increases the efficiency of concentration by collecting off-axis rays that would else be lost and directing them toward the exit opening. These rays reflect several times within the CPC and after such multiple bounces reach the exit. The rays arriving at an angle larger than the acceptance angle do enter the concentrator, but after bouncing several times off its walls they are lost, i.e., they do not reach the exit opening.

The total length of the CPC concentrator is

$$L = \frac{r(1 + \sin\theta)\cos\theta}{\sin^2\theta} \tag{2.5}$$

Obviously, a CPC can be truncated and in real situations often is. This will result in a lower concentration ratio, but the overall length might be more acceptable for implementations.

There is a relation between the radii of the entry and exit openings

$$R = \frac{r}{\sin\theta} \tag{2.6}$$

The equation defining the surface of a CPC in a radial system in which the CPC axis is in z-direction while the radial distance is described as the coordinate ρ is

$$\begin{aligned} 2r\rho\,(1 + \sin\theta)^2 + (\rho\cos\theta + z\sin\theta)^2 \\ = 2rz\cos\theta(2 + \sin\,\theta)^2 + r^2(1 + \sin\theta)(3 + \sin\theta) \end{aligned} \tag{2.7}$$

A special case of the rotational Winston collector is a CPC trough. Both 2D and 3D CPC structures are ideal concentrators. 2D CPC is convenient if the target travels along a line since then there is no need for its tracking.

One possible improvement of the CPC is to fill it with dielectric having a refractive index n. This shortens the structure and increases the acceptance angle. At the same time it uses total internal reflection which ensures a higher reflection than from metal surrounded by air. The maximum concentration ratio in this case is larger by a factor of n^2. Another useful modification is to utilize a two-stage CPC where the exit opening of the first stage CPC coincides with the entry opening of the second stage.

2.4.2 Parabolic Reflector

Geometrically, a 3D parabolic reflector represents a truncated paraboloid, as shown in Fig. 2.8. Parabolic reflectors are actually imaging optical elements, used in various optical systems. For instance, a parabolic reflector is often combined with a diverging mirror that directs light beams through the exit aperture. This is the well-known Cassegrain configuration, the standard design for reflector telescopes.

Fig. 2.8 Schematic presentation of a parabolic reflector. The parabola is cut off in the focal point

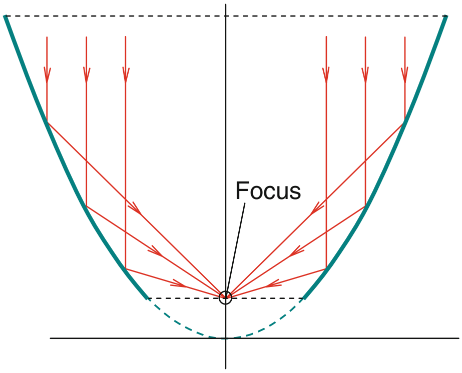
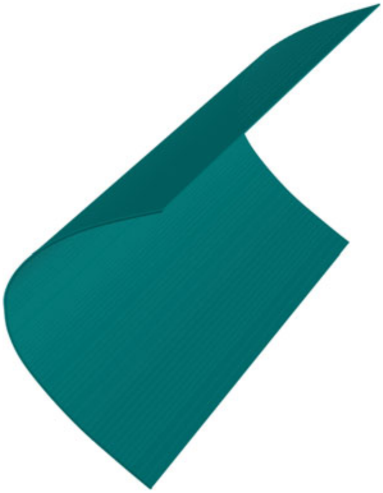


Fig. 2.9 Parabolic trough



The microfabricated version of parabolic reflector is used in solar energy concentration [130].

Contrary to the CPC, a parabolic concentrator is a high-concentrating device, but is more sensitive to scattering of incident radiation than the CPC. A problem with parabolic concentrators is that their total transmission is much smaller than that of CPC and reaches about 0.60.

A 2D implementation of the full paraboloid geometry is the parabolic trough [131], Fig. 2.9. Although a full three-dimensional paraboloid is more energy efficient than a trough, 2D implementation is less complex and can be useful in situations where there is a linear array of photodetectors.

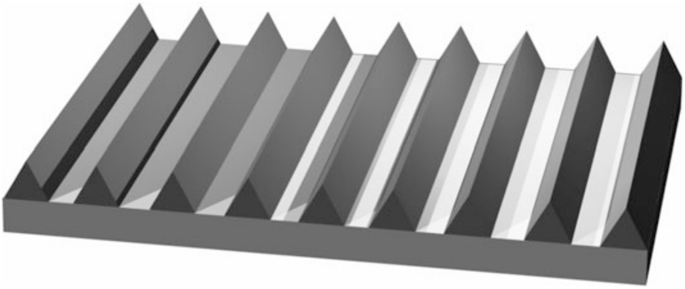


Fig. 2.10 A V-trough for reflective nonimaging concentrators

2.4.3 Conic Concentrator

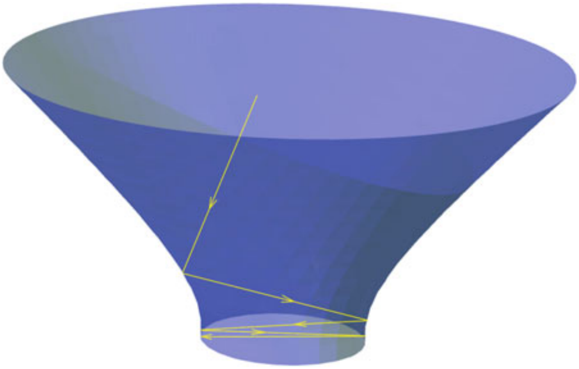
The conic concentrator is actually a truncated cone with reflective sides. It is not an ideal concentrator, and its transmission reaches about 0.8. Its 2D version (with a triangular profile) is called a V-trough, Fig. 2.10.

2.4.4 Hyperbolic (Trumpet) Geometry

A geometry for reflective concentrators is the trumpet concentrator [128], Fig. 2.11. This is actually a hyperboloid of revolution. Besides the CPC it is denoted as one of the "maximum collection efficiency" nonimaging concentrators. It is characterized by multitudinous reflections near the exit aperture.

Trumpet concentrators are ideal concentrators both in 2D and 3D case. They also offer the advantage of concentrating not only internal but also external rays (those latter are concentrated to an annular shape around the exit aperture). Trumpet concentrators are characterized by a small material expenditure. External rays are not usable for detector matrices.

Fig. 2.11 'Trumpet' concentrator



2.4.5 Other Reflector Geometries

There is a plethora of reflector geometries for nonimaging concentrator optics. Among them are elliptic and compound elliptic concentrators (CEC) as well as high-flux two-stage optical designs based on tailored edge-ray concentrators (ordinary and compound). They also include different lens-mirror combinations, utilizing reflection (X), refraction (R), and total internal reflection (I), for instance XR, RX, XX, XRI, etc. [128]. Complex forms for ultra-high concentrations of input radiation are obtained by using different algorithms, for instance global optimization procedures.

2.4.6 Fresnel Reflector

The idea of the Fresnel reflector is to divide a curved reflector into segments and to move these segments to a single joint plane [132, 133]. Thus one is able to effectively reduce a 3D object to a thin (quasi-2D) object while keeping the identical optical path to that of the 3D object. For instance, if one starts a 3D body of revolution like, for instance, a parabolic reflective concentrator or a Winston cone, this body can be divided into a number of annular segments which are subsequently placed concentrically on a flat surface. This process is illustrated in Fig. 2.12a, where a curved body is cut into segments and placed on a surface. An incident beam reflects under the same angles as it would from the full 3D body.

The surface of the original body of revolution can be turned into a Fresnel reflector by being cut into annular segments with the reflective area identical with the original, i.e., with curvilinear surfaces, Fig. 2.12b. It can be also approximated with blazed geometry, Fig. 2.12c, where the cross-section of the annular segments is fabricated in triangular form, with the internal side of the triangle following the original form as accurately as possible.

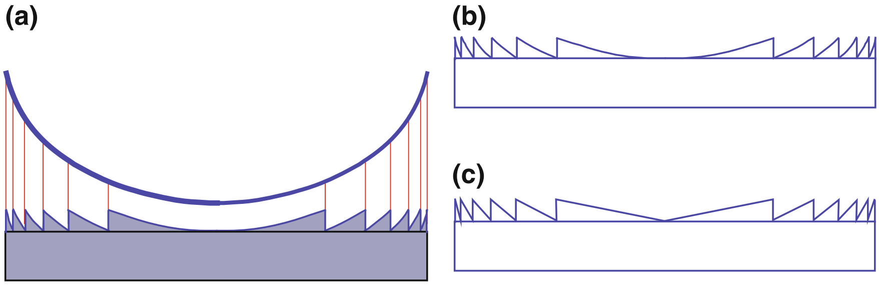


Fig. 2.12 Fresnel reflector. **a** Method of cutting reflective surface. **b** Reflective surface represented by in-plane curvilinear annular segments. **c** Reflective surface approximated by flat (blazed) segments with triangular profile

Fresnel reflectors are much more compact and convenient than full 3D bodies of revolution or 2D troughs. They can be both of imaging and nonimaging type, depending on the body of revolution (or 2D trough) they approximate.

2.5 Diffractive Optical Concentrators

2.5.1 Definitions

We start this section with some definitions. First, we define a diffractive optical element (DOE) as a passive optical element, its properties being either stationary or switchable, with operation based on diffraction or interference of light beams passing through it or being reflected from it. This is a very broad definition and includes a large number of specific subclasses.

Any DOE is basically an optical grating structure, either surface- or volume-type. The grating lattice is defined by some kind of modulation of the complex refractive index of the grating material.

A binary *DOE* is an optical grating element which is defined by a finite number of discrete levels in the values of its optical constants. The smallest value of this number is 2. Due to technological issues, this number is in general case equal to 2^N, where N is the number of technology steps required for its fabrication, typically photolithography steps. Some authors actually define binary optics elements as those passive optical elements which are fabricated using VLSI technology.

A grayscale *DOE* is a generalization of a binary element and represents those cases when N is large enough to be regarded as infinite.

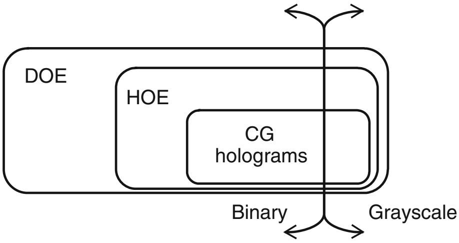
If a DOE is fabricated using holography, it is called a holographic optical element (HOE). The terms HOE and DOE are sometimes used interchangeably.

Any diffractive optical element may be computer-generated (CGH, computer-generated hologram) or designed directly using a physical interference pattern, either from some object or generated in an interferometer.

The family of diffractive optical elements is schematically shown in Fig. 2.13.

In dependence on which part of the complex refractive index was used to modulate the diffractive lattice, a DOE may be of amplitude type (the imaginary or absorptive part of the refractive index was used for modulation, i.e., the grating is

Fig. 2.13 The family of diffractive optical elements



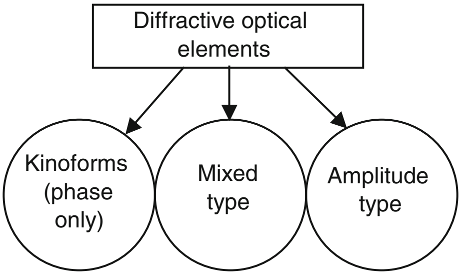


Fig. 2.14 Classification of DOE in dependence on the mechanism of refractive index modulation

built from transmissive and opaque parts, and thus only the amplitude of the incident beam is changed), phase type (only the real (refractive) part of the refractive index is modulated, thus only the beam phase is changed) and mixed (the material used changes both refraction and absorption of the incident beam). The diffractive optical elements influencing only the phase of the incident beam are kinoforms.

Figure 2.14 shows the division of DOE to classes in dependence on refractive index modulation.

Because DOEs function by breaking the incident light wavefront into wavelets and restructuring it, they offer a larger degree of freedom in designing their functions. Besides replacing most of the conventional optical elements, DOEs have enabled the development of some products that could not be produced with previous optical technologies. In addition to that, their area can be larger, even much larger than that of the conventional optical elements, they are much thinner and their cost is lower. They also offer an enormously wider choice of materials for their fabrication, providing the user with a greater flexibility in selecting the material for a particular application.

A diffractive optical element can be either transmissive or reflecting. It may be polarization sensitive or it can be insensitive to it.

One of the most important features of DOEs is their angular and wavelength sensitivity or dispersivity. This means that their focal length and aberrations vary with wavelength and incident angle.

The properties of some DOEs are constant within a narrow wavelength band, e.g., 10–20 nm, which may be a serious shortcoming for some applications. However, there are solutions for multicolor applications, among them being the use of DOEs that widen the operating frequency band by integrating in themselves several separate narrow-band elements, each covering a part of the desired spectrum. Also, some DOEs have wider wavelength ranges.

Since almost any material may be used for DOEs, the available media include electrochromic, liquid crystal and other recording materials. Thus DOEs can be easily switched between two or more desired functions, i.e., recorded and erased and generally controlled in time by the user.

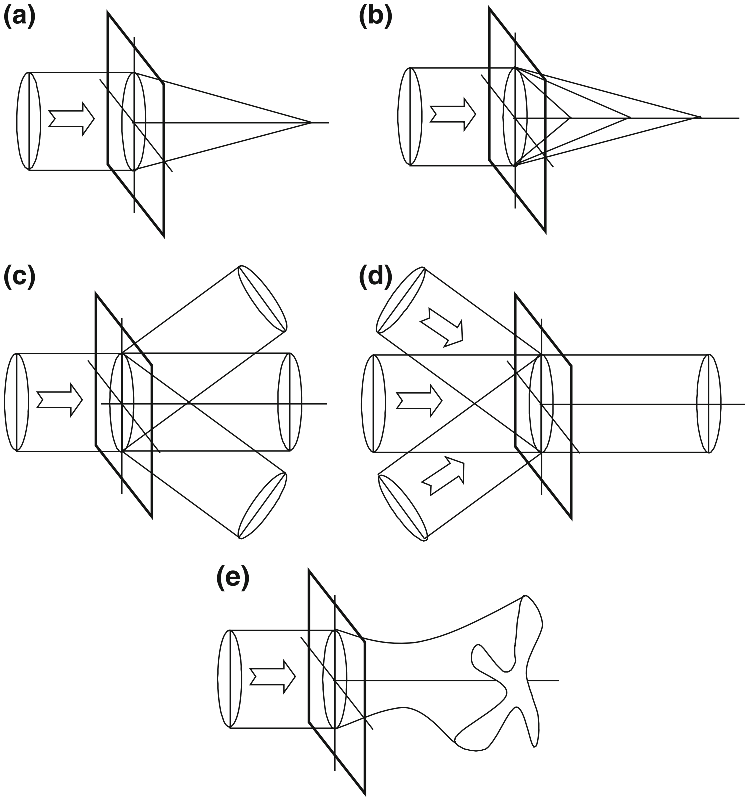


Fig. 2.15 DOE functions of interest in photodetection. **a** Simple focusing; **b** focusing with variable focal length; **c** beam splitting; **d** beam combining; **e** beam shaping

Diffractive elements are very lightweight, as they are typically formed in thin films of a few micrometer thickness only (this is valid even for the so-called "volume structures" where a thickness of only a few wavelengths usually suffices.) They can be completely flat even if they perform functions of conventional elements with curved surfaces. On the other hand, a DOE can be fabricated on any arbitrary shape of the substrate. Their fabrication is cheap in comparison to conventional optics (typically a master element is fabricated and its replicas are created by some mass production methods).

Figure 2.15 shows some functions achievable by the use of diffractive optical elements that are of interest for photodetection. The principal function of interest is focusing or light concentration. This may be done in a way similar to conventional lenses (Fig. 2.15a), but one of the possibilities offered by DOEs is also multiple focusing, i.e., simultaneous focusing of different beams to several foci at once with a single DOE, in dependence on the beam properties or on external controls. These foci may be along the same axis (only focal length is changed) (Fig. 2.15b) or separated in space (both focal length and focus position are changed)—beam-splitting with simultaneous focusing. Naturally, DOEs may be utilized for beam-splitting without focusing (Fig. 2.15c). The selectivity of a DOE may be designed to depend on any desired property of the incident waves (e.g., their wavelength, polarization, spatial angle of incidence, etc.). DOEs in photodetection may be utilized as beam combiners (Fig. 2.15e).

Practically, it may be said that several different diffractive optical elements can coexist on the same substrate without interfering with one another. They can perform the same function, either selectively, as described above, or implement completely different functions. For example, a single DOE can be simultaneously used as a lens, beam splitter and spectral filter.

A DOE can be used for beam shaping, i.e., it can transform an incident wave from a given initial form to a desired final form (Fig. 2.15e), for example from planar to Gaussian distribution or vice versa. It can serve as a generator of arbitrary aspheric waves, or a most general wavefront transformer. This includes both spatial redistribution and phase correction and may incorporate generation of an interferogram of existing or nonexistent object, sampling of wavefronts, etc.

Finally, DOEs can serve as various optical spectral filters, including low-, high-, or bandpass filters, spectral transformers, notch-filters, etc. All of these features can be used to enhance photodetectors.

2.5.2 Fresnel Lenses

Probably the most important and definitely the simplest DOE for optical concentration in photodetection is the Fresnel Zone Plate (FZP) [134]. An FZP lens represents a set of concentric rings with their width decreasing toward the edge. Light is diffracted from these rings and the diffraction maximum is in the focal point of the lens. The alternating transparent and opaque rings are calculated to result in adding the electromagnetic waves with the same sign, and blocks those with the opposite sign. The resulting function is similar to that of a conventional convex lens and the structure works as a light concentrator. The intensity of the electric field in the focal plane is linearly proportional to the number of the zones, i.e., the number of transparent rings. The maximum diffraction efficiency (defined as the ratio between the incident and the transmitted radiation power) of the FZP does not exceed 10 %, because approximately one half of the radiation is stopped at the opaque rings, while the phase distribution of the other half is sinusoidal.

A FZP is a 2D form and a large number of annular lenslets present in this form are arrayed on a single optical surface, thus replacing a thick refractive optical lens by an essentially planar structure. Figure 2.16 represents a typical FZP. Instead of having sharp edges, the transitions between the opaque and transparent zones are more gradual, which is a situation often met in reality.

The optimum diameter of a Fresnel zone for a case of a planar incident wave is calculated as [135]

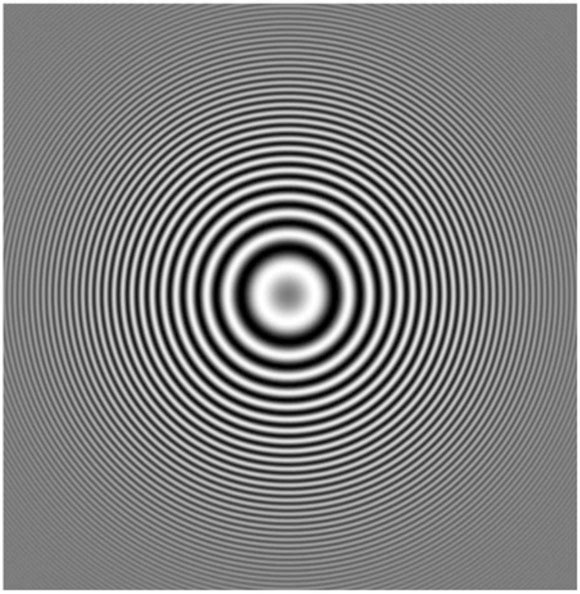


Fig. 2.16 Fresnel zone plate

$$D_m = \sqrt{m f_F \lambda + \frac{m^2 \lambda^2}{4}}, \tag{2.8}$$

where λ is the wavelength, f_F focal length, and m is the number of the ring (counted from the center of the FZP.)

The irradiation distribution in the focal plane for a binary amplitude FZP is calculated by the Fresnel-Kirchhoff integral, performing integration across all of the transparent rings $S_1,\ldots,S_n$ [135]

$$E_p = -\frac{E_{0i}}{\lambda} \oiint\limits_{S_1,\ldots,S_n} \frac{e^{i\frac{2\pi}{\lambda}r}}{r} \frac{1-\cos\left[\overrightarrow{n}\,\overrightarrow{r}\right]}{2} ds \tag{2.9}$$

where E_p is the electric field in the observed point of the focal plane, E_0 is the amplitude of the electric field of the plane wave, $\vec{r}$ is the radius vector of the surface element dS FZP, $\cos[\vec{n}, \vec{r}]$ is the cosine of the angle between the normal to the FZP and the radius vector of the point P observed from the dS, S_n is the area of the n-th transparent zone.

Figure 2.17 shows the distribution of irradiation in the focal plane, calculated numerically using (2.9). The irradiation power drastically decreases and the focus diameter increases if the number of the zones is smaller.

We further give a connection between the dimensions of a binary amplitude FZP and the active area of an IR detector. For the sake of simplicity, we assume that the

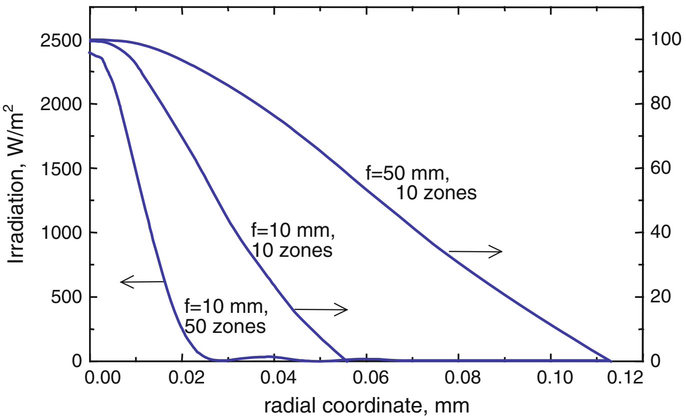


Fig. 2.17 Distribution of irradiation intensity in the focus of a Fresnel zone plate for various numbers of zones and different focal lengths

active area is circular and that a planar monochromatic wave illuminates the FZP. The increase of the electric area of the detector is

$$\frac{A_0}{A_e} = \eta_F \frac{\pi R_F^2}{\pi r_d^2} \tag{2.10}$$

where R_F is the radius of the outer zone of the FZP, FZP, r_d is the radius of our idealized detector, and η_F is the efficiency of the FZP (the ratio between the incident and the transmitted radiation power), here equal to 0.1.

The effective area can be increased either by increasing the FZP radius or by decreasing the photodetector-active area. The maximum diameter of the diffractive lens is limited by the properties of the photomask and by the dimensions of the housing. The minimum diameter of the idealized detector is equal to the diameter of the focus. In the above case, when the focal length is 10 mm, and the FZP diameter calculated according to (2.2) is 5 mm (50 zones), as shown in Fig. 2.18, the optimum detector diameter is about 20 μm. Such a detector coupled with the binary amplitude FZP theoretically receives about 2,000 times more radiation than the same device without a concentrator. For a detector with an active area diameter of 250 μm, the effective improvement achieved by the same diffractive lens is 40 times.

In the case of a detector for the detection of CO_2 laser radiation (10.6 μm) with a FZP diameter of 5 mm, the total number of Fresnel zones 50, and the distance to the active area 10 mm, the focus diameter is 650 μm.

The FZP is the simplest and the least efficient planar diffractive optical concentrator of Fresnel type. Other types are shown in Fig. 2.19.

The simplest and the least efficient binary DOE is the mentioned Fresnel zone plate (Fig. 2.19a), containing only alternating opaque and transparent rings (i.e., an

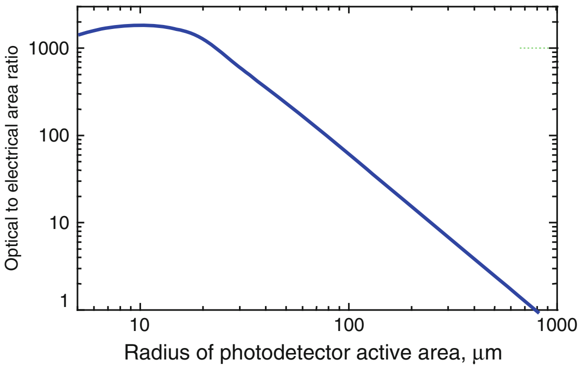


Fig. 2.18 Ratio of optical and electrical area in dependence on the active area radius of a photodetector for a binary amplitude FZP with 50 zones, intended for the detection of IR radiation at 10.6 μm and with a focal length of 10 mm

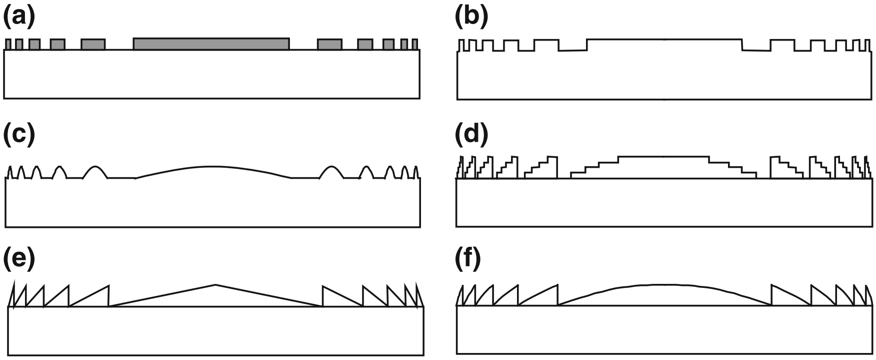


Fig. 2.19 Various types of DOE lenses. **a** Fresnel zone plate; **b** Fresnel phase plate; **c** Fresnel phase plate with "soft" edges; **d** binary multilevel phase FZP; **e** blazed Fresnel lens; **f** Fresnel lens (continual kinoform lens, equivalent to a binary DOE lens with an infinite number of levels)

amplitude-modulated element). Similar but more efficient is the single-level Fresnel phase plate (Fig. 2.19b), where the opaque rings are replaced by phase-shifting rings introducing a phase change of π/2 (a kinoform-type DOE). Instead of losing 50 % of the incident light intensity, one utilizes all illumination.

In reality, the Fresnel phase plates are often with gradually changing edges, Fig. 2.19c.

The next level of sophistication is the multilevel binary Fresnel lens (Fig. 2.19d). It represents a grayscale-DOE with a finite number of steps (due to the nature of photolithographic process used in its fabrication, it will have 2^N levels, where N is the number of the photolithographic steps.

The diffraction efficiency of a multilevel Fresnel phase plate (the ratio between the incident and the transmitted optical power η_F is calculated as [136]

$$\eta_F = \sin^2\left(\frac{\pi/M}{(\pi/M)^2}\right), \tag{2.11}$$

where $M = 2^m$, while m is the number of levels. For the case shown in Fig. 2.19b ($m = 1$), the efficiency is 41 %, and 99 % for $m = 4$.

The next level is the blazed planar lens where the steps are replaced by sloped planes (Fig. 2.19e).

The most sophisticated is the full Fresnel lens, actually a curved-surface kinoform (Fig. 2.19f), which corresponds to a multilevel binary Fresnel plate where N tends to infinity, but also corresponds to a planarized conventional refractive lens. In other words, this Fresnel lens would be obtained if a conventional lens was cut into rings and if each ring was then thinned to equal thickness by removing from its basis all the excess material that does not contribute to focusing.

2.5.3 Holographic Optical Elements

A holographic optical element (HOE) is a diffractive optical element either designed by computer or used for image forming by holography. In other words, a DOE can be denoted as a holographic optical element if fabricated by holography or intended to be used as a hologram. As an example, a DOE fabricated by ruling using a diamond cutter is not a HOE, but an identical DOE fabricated holographically belongs to the HOE. The design of holograms using numerical methods on a computer is called Computer-generated Holography (CGH).

Holography is defined as a method of recording and reconstructing wavefronts, and optical holography is connected with the light in visible, ultraviolet, and infrared spectrum. A hologram is a material recording of the wavefront phase and intensity distribution of a light beam. Thus it is actually an interferometric pattern created either by an interferometer, a wavefront from a real object interfering with another coherent beam, a wavefront from another holographic recording or a computer-generated interferometric pattern corresponding to an existing or even a nonexistent object. Obviously, holographic fabrication may be used to manufacture any type of diffractive optical element.

Holograms can be classified in a manner similar to DOEs as amplitude (absorption-type) HOE, phase only (kinoform), and mixed ones. Another classification is based on their thickness. They are thin (or planar) if their total thickness is not much larger than their fringe spacing (i.e., the operating wavelength). If it is, they are thick (volume, Bragg-type) holograms.

A volume hologram may be regarded as a superposition of three gratings recorded in the depth of the medium, each satisfying the Bragg law. A volume hologram produces maximum change in refractive index and/or absorption index and offers the largest diffraction efficiencies. It reconstructs the virtual image at the original position of the object if the reconstruction beam exactly coincides with the reference beam. The conjugate image and higher order diffractions are absent in a volume hologram.

Focal length and wavefront shape of an output beam from a HOE do not depend on the refractive index of the utilized material, although it influences, e.g., the angular dependence of some of its parameters. HOEs are used if it is not practical or is impossible to use refractive optics in a given application [137]. Among the advantages of HOEs are their reconfigurability, the possibility to obtain several different functions in a single element (see Fig. 2.2), the possibility to use beam-splitting to focus radiation on a larger number of spatially separated positions (as in detector arrays), the beam-shaping possibilities to adjust the focused beam to the shape of a detector element or a detector array, the possibility to separate different modes of incident radiation to different detector elements, etc.

Among HOEs an important position belongs to kinoforms, thin-phase holograms with a surface microrelief corresponding to the phase modulation of a given object [138, 139] (i.e., without any data on the radiation amplitude.) Kinoform-based HOEs can operate with incoherent light, i.e., it may be designed for a range of operating wavelengths and allows a relatively wide field of view. Kinoform may be transferred from standard holographic materials and replicated in MWIR and LWIR materials using a number of microfabrication procedures, e.g., reactive-ion etching [140].

HOEs may be regarded as a generalization of Fresnel lenses and their definitions partly overlap—for instance, the kinoform structure shown in Fig. 2.2e is a typical HOE. We may say that Fresnel lenses are a subgroup of holographic optical elements.

The most important parameter of HOEs for infrared detection is η, diffraction efficiency, defined as the ratio between the incident optical flux and the flux forming its image. It is predominantly dependent on absorption losses, i.e., on the absorption coefficient of the HOE material within the given operating wavelength range, but also on the incident angle. Diffraction efficiency of a general HOE is calculated using the coupled-wave theory [134]. The maximum η is obtained for transmission-type phase HOEs (e.g., kinoforms, $\eta \approx 1$), while it is almost two orders of magnitude lower for transmission-type amplitude HOEs ($\eta \approx 0.037$). The second parameter usually applied to describe the behavior of a HOE, signal-to-noise ratio, is of much lesser importance for optical concentrators for photodetectors.

Holographic optical elements with desired characteristics are typically designed using the methods of computer-generated holography (CGH) [141]. The most often used CGHs are the Fourier holograms [142]. Two predominant methods for the calculation of CGHs are Lohman's [143] and Lee's [144]. Holograms are computer generated using the FFT algorithm (Fast Fourier Transform) [145]. The reconstructed image is determined by the inverse FFT. In real optical systems, the Fourier transform of an optical signal is obtained by using refractive lenses.

In Fresnel-type HOEs, the light distribution in the hologram plane is the Fresnel diffractive image of the object. For radially symmetrical kinoforms (rings and circles), 2D Fourier and Fresnel integrals are reduced to 1D Hankel's integrals [146].

During the last years, non-Fourier CGHs are increasingly used [147]. They are used without any additional optics and allow the design of holograms of arbitrary objects, including those not existing in reality, as well as the design of full 3D holograms. Their shortcomings is that they are extremely demanding for CPU resources, especially when calculating reconstructed images. Usually some acceleration methods are used, e.g., the Monte-Carlo method [146] or smaller 2D arrays are generated, e.g., 512 × 512 pixels, and subsequently repeated.

One point is important for the use of HOEs in infrared detection. Standard holographic materials today are photosensitive mostly in the short-wavelength part of the visible spectrum. Since they must be used for infrared radiation, this could cause large aberrations, as well as the breaking of the Bragg condition, especially if the incident wave is not planar. This is the reason why holographic optics for larger wavelengths are typically designed utilizing some of the iterative CGH methods, e.g., Gerchberg–Saxton's algorithm [122]. In this manner, one obtains a holographic optical element diffraction-limited in a desired wavelength range [148]. To this purpose, a Fourier transform is generated or some other method is used to calculate the CGH to be fabricated using standard procedures. Then the image is reconstructed at the operating wavelength (MWIR or LWIR). Finally, the function of the reconstructed image aberration is minimized (a variant of the so-called method of conditional gradient) [122]. Due to a slow convergence and the effect of stagnation, some methods of improvement of these iterative procedures are introduced, e.g., reset of the complex amplitude in each iterative step [149] or the use of adaptive algorithms [122].

DOE lenses for HgCdTe detector arrays were described within the context of HgCdTe detector focal plane arrays [150]. Diffractive (8–16-phase-level) Ge microlenses reducing the detector size in focal plane arrays by 3–6 times were described in [151].

2.6 Antireflection Structures

The increase of incident optical flux into an optical element by decreasing reflection from the incident surface is one of the oldest problems in optics (e.g., [87]). The customary approach is to form some kind of antireflection (AR) structure directly on the surface of the optical element. Antireflection structures may be denoted as thin homogeneous or inhomogeneous layers deposited on the incident optical surface or formed directly in it with the goal to use destructive interference or diffraction to maximize the amount of flux penetrating the interior of the element. An optimum antireflection structure for a given application furnishes the minimum of reflection for given optical properties of the element and the ambient within a given spectral range.

The use of AR structures in the case of IR detectors is even more important than in the visible spectrum, because materials used for the fabrication of both active regions of the detectors and optical concentrators usually have large real parts of refractive index (typically 3.5–4, for instance in the case of InSb for MWIR, HgCdTe for LWIR and Ge for optical windows and concentrators.) Antireflection structures are usually deposited directly on the active area of the detector, on the collimating optics and on the optical windows.

In our consideration, we assume that the AR structure is nonabsorbing or weakly absorbing, which is valid even for the AR structures incorporating narrow-bandgap semiconductors because of their very small total thickness within a given AR structure. The necessary data for the design of an optimized AR structure are the required spectral range of minimized reflection, the real part of the refractive index of the substrate and the ambient, as well as the minimum and maximum limits of the refractive index for the AR structure. Sometimes the maximum optical thickness of the whole AR structure is also given.

Two main groups of antireflection structures are single- or multilayer AR layers composed of dielectric films (interference films) [152] and single or multiple periodical diffractive structure of either amplitude of phase type. Multiple AR structures may have quarterwave or quarterwave/half-wave periodicity or they may be aperiodic [153].

2.6.1 Antireflection Interference Films

An optically thin film may be defined as a plane-parallel structure with a thickness of the order of the wavelength of light, characterized by its complex refractive index and thickness. The simplest and cheapest, but the least effective approach to the minimization of reflections on the incident optical surface is to deposit one or several optically thin films on it, with a thickness and refractive index value optimized to furnish maximum transmission. The method is based on destructive interference between the strata within the AR thin film. Such interference-based dielectric films [152] or interference coatings were chronologically the oldest structures used to minimize reflection on the incident optical surface. They consist of layers whose optical thickness (the product of their real thickness and refractive index) is equal to the targeted quarter-wavelength multiplied by an odd integer. This method was conceived in the early 1950s [87]. Usually, the refractive index is alternating between neighboring layers, thus forming layer pairs, while the optical thickness is kept constant. During the last few decades, optimized AR films are also produced by varying both the optical thickness of layers and their refractive index.

Antireflection interference thin-film coatings based on homogeneous strata may be grouped as [154]

(a) single layer,
(b) multilayer, periodic,
(c) multiple, quasiperiodic,
(d) multiple, aperiodic (for instance step-down profile), and
(e) multiple, random.

The simplest AR coatings are the single-layer ones. They consist of a single homogeneous layer used for most of the low-end applications. Currently, they still comprise the majority of all existing AR coatings, since their performance is satisfactory if a transmission of the order of 95 % fulfills the posed requirements. Theoretically, there is a single wavelength within a given range for which a single AR layer furnishes 100 % transmission.

For single-layer quarter-wavelength homogeneous antireflection film, the condition of reflection minimum can be calculated as

$$n_1 = \sqrt{n_0 n_s}, \tag{2.12}$$

where the refractive index of the single homogeneous AR film is n_1, that of the substrate n_s and of the surrounding medium n_0. The optical thickness of the layer should be

$$d_1 = \frac{\lambda}{4n_1}. \tag{2.13}$$

Multilayer structures offer much better performance and vastly superior possibilities for optimization. Among their main advantages are the possibility to obtain minimal reflection in a wider spectral range and facilitated choice of convenient materials and technologies.

The condition of reflection minimum for a double-layer quarter-wavelength antireflection film made from homogeneous strata is

$$n_1 n_2 = \sqrt{\frac{n_s}{n_0}} \tag{2.14}$$

where the refractive indices of the first and the second film within the AR coating are n_1, and n_2.

The condition of reflection minimum for a three-layer quarter-wavelength antireflective film with homogeneous strata is

$$\frac{n_1 n_3}{n_2} = \sqrt{\frac{n_s}{n_0}} \tag{2.15}$$

where the refractive indices of the first, second, and third film are n_1, n_2, and n_3, respectively.

With an increase of the number of layers, the spectral bandwidth of the reflectivity minimum becomes wider and the mean deviation from the zero reflectivity lower.

Figure 2.20 shows a schematic presentation of a generalized multilayer AR film with a total of m layers. The refractive index and the thickness of each layer can be different.

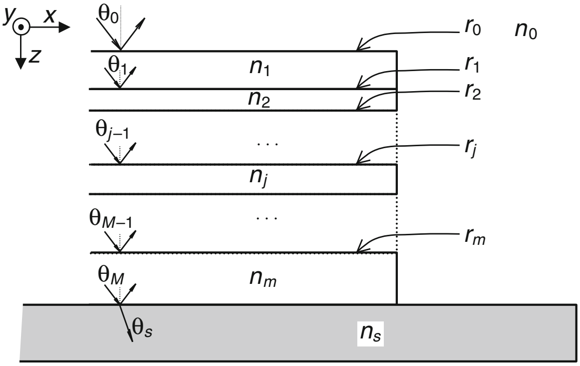


Fig. 2.20 Multilayer antireflection film consisting of discrete and homogeneous all-dielectric strata

Reflection and transmission coefficients for multiple interference antireflection layers may be calculated in the following manner. For each j-th layer, we write the expression for Fresnel amplitude reflection coefficient r_j^* for the next m–j layers located behind it. This expression features the reflection at the surface of this layer r_j, its phase φ_j and the amplitude reflection coefficient of all the remaining m–j–1 layers. By writing successively in the same manner the expressions for the remaining layers, we obtain a recursive system with an infinite number of solutions

$$r_j^* = \frac{r_j e^{i\varphi j} + r_{j+1}^* e^{-i\varphi_{j+1}}}{e^{i\varphi_j} + r_j r_{j+1}^* e^{-i\varphi_j}}, \tag{2.16}$$

$$r_j = \frac{n_{j-1} - n_j}{n_{j-1} + n_j}, \tag{2.17}$$

$$j = 1, 2, \ldots, j, \ldots, m.$$

Here m denotes the total number of layers, n_j is the real part of refractive index for the j-th layer. The layers are all-dielectric, i.e., the absorption of the AR film is neglected. The amplitude reflection coefficient of the whole film is further made equal to zero (the condition for the reflection minimum) and then the parameters of the whole multilayer structure are numerically optimized.

For given spectral limits, the best results are obtained if using discrete layers with homogeneous composition and with a quarter-wavelength optical thickness [155],

$$d_1 n_1 = d_2 n_2 = \cdots = d_j n_j = \cdots = d_m n_m = \frac{\lambda}{4} \tag{2.18}$$

For a substrate with a refractive index larger than that of its ambient it is simultaneously necessary that

$$n_0 < n_1 < \cdots < n_j < \cdots < n_s \tag{2.19}$$

(step-down refractive index) [155].

The thickness of the j-th layer is

$$d_j = \frac{d}{\sum\limits_{\substack{k=0 \\ j\neq k}}^{m} \left(\frac{n_j}{n_k} + 1\right)}. \tag{2.20}$$

For a single AR layer, the above condition reduces to $d = \lambda/4n_1$ and $n_1 = n_S^{1/2}$, assuming that the incident medium is air.

For multilayer quarter-wavelength antireflection films with homogeneous strata (step-down refractive index film) [156, 157], the amplitude reflection coefficient determined for the whole structure in the aforementioned manner is made equal to zero (the condition of reflection minimum) and numerical optimization is further performed to obtain the strata parameters. In that case, the above-mentioned step-down condition is applied that refractive index decreases in steps from the n_s value at the substrate surface to the n_0 value of the surrounding medium [155].

A problem with quarter-wavelength layers is to find materials with convenient refractive index, since material with the required refractive index value may not even exist. There are several possible approaches to this problem [158]. One is to utilize a combination of two material furnishing an effective refractive index near the required value. Another way is to deposit two homogeneous layers whose transfer matrix is equal to the matrix of the desired material (the Herpin equivalent-index concept) [159]. Third method is to use AR layers aperiodic both regarding the layer thickness and their composition. It was shown that a combination of an arbitrary number of different materials can be mathematically reduced to a combination of two materials only, for instance by utilizing the equivalent-index concept. This fact may be used to simplify the design of AR layers.

The problem to find an optimal AR film in the most general case when both thickness and refractive index may be arbitrarily changed reduces to an optimization procedure in a multidimensional parametric space. It is assumed [160] that for a given spectral range and a given interval of values of refractive index there must be a single optimal generalized AR film structure, i.e., there must be a global solution of the problem, giving the lowest reflection coefficient in the whole range.

A method to solve the problem is to determine in the Fourier space the connection between the logarithm of refractive index values and the amplitude reflection and transmission coefficients, represented as complex wavelength-dependent functions. The global minimum of thus obtained dependence is then determined. The solution is an inhomogeneous layer, further transformed into a two-material system and subsequently subjected to a new procedure of fine optimization.

Among a number of approaches to the calculation of the global minimum, we mention the needle optimization and the method of quadratic programming [153]. The needle optimization obtained its name because it is based on a procedure of introducing a Dirac function ("needle") into the refractive index profile and thus determining the best position to introduce a new layer; fine optimization is performed afterward. The procedure is repeated until no new layers can be introduced.

The method of quadratic programming starts from a quality function equal to the integral of the ratio $R(\lambda)/[1 - R(\lambda)]$ over wavelength from the beginning to the end of the optimization interval. The solution here is also an inhomogeneous layer transformed subsequently into a two-material one and finely optimized [158].

Different approaches to the calculation of the global reflection minimum of a generalized AR film furnish very similar solutions, independently on the algorithm used and the initial conditions, which points to an existence of a unique solution for the given conditions. The maximums of refractive index in an optimized AR film are grouped. This is a consequence of the fact that complex filters are basically a combination of two interference filters with different periods [153]. A combined AR film profile is a result of the beating of their periods. The long-wavelength filter causes the mentioned grouping, determines its period, and creates a high-reflection peak at the long-wavelength boundary of the AR range. The short-wavelength filter determines the fine structure of groups and defines a reflection peak at the opposite end of the AR range. The reflection peaks at the boundaries of thus designed AR layers are very high, reaching almost 100 %.

The main disadvantage of all-dielectric interference-based AR films is that due to their very nature their performance sharply drops for oblique angles, i.e., only small deviations of the incident angle are permitted from the value for which a coating was designed. A typical AR interference coating is designed for normal incidence of light. This can be somewhat improved if material pairs with a large refractive index ratio are used. Other problems include technological issues, for instance cohesion problems and thermal matching between the substrate and the deposited layers. A design problem is a limited choice of materials with suitable refractive values of refractive index. This can be alleviated by the use of porous materials.

The most often used materials for the fabrication of AR layers in the MWIR and LWIR range are ZnS, MgF_2, and similar ones. Magnetron sputtering is very often used, reactive or nonreactive, either dc or RF. It gives layers with optical quality while at the same time offering the possibility to control their thickness very accurately. Among the methods used to deposit thin single or multiple coatings with antireflective properties is the deposition from solution and further curing, baking, and chemical processing. Various kinds of evaporation are used, both reactive and nonreactive. A method offering less porous coatings is ion-assisted deposition. One of the methods is to deposit first a thin metal layer and then to convert it to oxide or hydroxide AR coating. Chemical vapor deposition may also be used.

As an illustration, Fig. 2.21 shows experimental reflection diagrams for a germanium optical concentrator for uncooled HgCdTe photoconductors with a single-layer zinc sulfide AR coating deposited by RF sputtering.

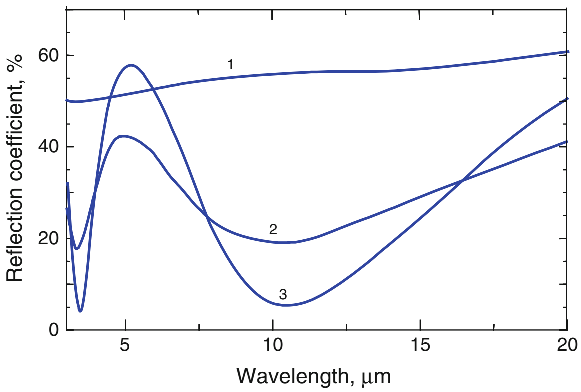


Fig. 2.21 Experimental spectral reflection for single-layer ZnS AR coatings. *1* Ge substrate 2 mm without AR coating. *2* ZnS layer 1.2 μm deposited on the top surface of the Ge substrate; *3* The same layer as in (*2*) deposited both on *top* and the *bottom* surface

2.6.2 Gradient-Index Single-Layer Dielectric Films

If the refractive index of a coating is reduced uniformly and continually within the AR coating from the substrate value to the surrounding medium value, excellent antireflection properties are obtained in a very wide wavelength range (e.g., [161, 162]). This is not surprising, since this situation can be represented as a step-down index profile with an infinite number of infinitesimally thin homogeneous strata. However, contrary to the conventional interference-based AR single- or multilayers, their operation is omnidirectional. They are polarization insensitive, i.e., their reflectivity is low for both transverse electric and transverse magnetic electromagnetic waves. A theory of gradient-index thin films can be found in [163].

Different functional dependences may be used for gradient-index AR layers. The simplest one is linear, but various polynomial dependences are met, for instance cubic or quintic. Sinusoidal dependence is also used. Of these, quintic dependence has been reported as the closest to optimum [164].

The fabrication of gradient-index materials with refractive index values ranging from that of the air to the values found in semiconductors presents significant technological challenge [152]. Optical materials with the values of refractive index close to 1 are difficult to produce. One of the possible approaches is to use porous materials whose effective refractive index nears that of the air. Xi et al. reported material with an effective refractive index of 1.05 [165] based on porous silica utilizing nanorods produced by oblique-angle vapor deposition. Minot used chemical etch-leach process to produce porous silica gradient films [166]. Electrochemical etch procedures for porous silicon structures furnishing controllable refractive index gradient was reported in [167]. Polymer-based porous gradient-index films using polystyrene-block-poly(methyl methacrylate) were described in [168].

2.6.3 Antireflective Diffractive Structures

Antireflective properties may be generated by forming a subwavelength surface relief diffractive structures on the detector-active area (ARS, antireflective structures [169]). Such structures are also denoted as high spatial frequency structures or subwavelength surface textures [119]. Antireflection structures generally have a twofold function in light management in photodetectors. Since effectively they represent gradient-index structures, they ensure minimum reflection at the surface. However, they also function as scatterers for light arriving from the inside of the active region and return the beams back, which also increases the optical path and ensures improved absorption.

Most of the ARS function as diffractive gratings of the zero-order. The optical behavior of subwavelength gratings can be satisfactorily approximated in a majority of situations by using the effective medium theory [170]. Figure 2.22 shows two main situations that can be encountered.

One of the presented structures is a monodispersion of subwavelength inclusions i (spheres) in dielectric host h, Fig. 2.22a. The other is polydispersion, Fig. 2.22b. The first situation can be described by the well-known Maxwell-Garnett model [171], the oldest effective medium model, obtained by the use of Clausius-Mossotti/Lorenz-Lorentz equation. The other case is polydispersion, described by the implicit Bruggeman expression [172, 173].

Diffractive ARS are usually calculated either using the effective medium approximation, or the rigorous coupled-wave analysis [169]. Direct numerical solution of Maxwell equations and the approach using the plane wave theory were also used [174].

The effective medium theory (EMT) or the homogenization method regards a structured diffractive medium as a homogeneous material with an effective refractive index equal to a weighted and spatially averaged value of the refractive index of both structural materials. This is a very old theory and it was considered even in the works of Lord Rayleigh near the end of the nineteenth century.

The simplest is the zeroth-order EMT, that considers the diffractive medium in static approximation (grating transmits radiation as an unstructured medium). The

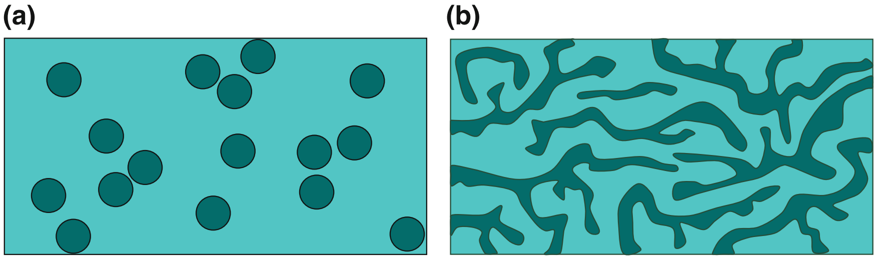


Fig. 2.22 Effective medium approach. **a** Spherical subwavelength inclusions as monodispersion in homogeneous host (Maxwell-Garnett model); **b** polydispersion (Bruggeman model)

condition of validity of the zeroth-order EMT is obtained according to the well-known Bragg equation of optical grating where the diffractive order is made equal to zero

$$\frac{\Lambda}{\lambda} < \frac{1}{\max\left(n_j, n_s\right) + n_1 \sin \theta_{\max}} \tag{2.21}$$

where $\theta_{\max}$ denotes the maximum incident angle. The condition of the existence of a zeroth-order transmitted wave is

$$\begin{aligned} n_j < n_s &\rightarrow 0 \leq \theta \leq \pi/2 \\ n_j > n_s &\rightarrow 0 \leq \theta_j \leq \arcsin\left(\frac{n_s}{n_j}\right) \end{aligned} \tag{2.22}$$

Thus we can calculate the angles of total reflection for the propagation through a subwavelength grating from an optically denser to an optically thinner medium. For instance, for HgCdTe this angle is 16.5°. Therefore the flux within the active area is additionally increased, because the returning beam is reflected from the incident surface back into the active region.

Based on the above diffraction equation, we introduce the miniaturization factor of an AR structure as $\Lambda/\lambda = [\beta(n_j + n_s)]^{-1}$. This factor should be maximized in order to simplify technological procedures for grating fabrication. In practical situations, the ratio between the grating step and the wavelength is about 10.

The zeroth-order EMT is not applicable for large refractive index modulations. Also, due to technological reasons, large patterns not satisfying the conditions of validity of zeroth-order EMT are used for ARS. In these cases, the second-order theory is used. It was developed in 1956 by Rytov for the case of magnetic and conductive materials [175–177]. In the case of ARS, one can utilize its form for dielectric media [178–180].

First, we define the grating vector (stratification vector) $\vec{K}$ as a normal to the boundary between two periods of a 1D optical grating. We utilize a coordinate system where the x-axis is parallel with this vector, while the y-axis is normal to it and to the incident surface. We use f_S to denote the fill factor equal to the ratio of the width of the part with the material having a permittivity ε_s to the total grating period Λ.

In the second-order EMT, the effective value of relative dielectric permittivity (square of effective refractive index) can be written as [176]

$$\begin{aligned} {\varepsilon_{\vec{E}\|\vec{K}}}^{(2)} &= {\varepsilon_{\vec{E}\|\vec{K}}}^{(0)} \left[1 + \frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2 f_S^2(1-f_s)\frac{\left(\varepsilon_s - \varepsilon_j\right)^2}{\varepsilon_0}{\varepsilon_{\vec{E}\perp\vec{K}}}^{(0)}\frac{{\varepsilon_{\vec{E}\|\vec{K}}}^{(0)}}{\varepsilon_s\varepsilon_i}\right] \\ {\varepsilon_{\vec{E}\perp\vec{K}}}^{(2)} &= {\varepsilon_{\vec{E}\perp\vec{K}}}^{(0)} \left[1 + \frac{\pi^2}{3}\left(\frac{\Lambda}{\lambda}\right)^2 f_S^2(1-f_s)\frac{\left(\varepsilon_s - \varepsilon_j\right)^2}{\varepsilon_0}\frac{1}{{\varepsilon_{\vec{E}\perp\vec{K}}}^{(0)}}\right]. \end{aligned} \tag{2.23}$$

The index (0) denotes the EMT approximation of the zeroth order

$$\begin{aligned}\varepsilon_{\vec{E}\perp\vec{K}}{}^{(0)} &= f_S\varepsilon_s + (1 - f_s)\varepsilon_j \\ \varepsilon_{\vec{E}||\vec{K}}{}^{(0)} &= \left[\frac{f_S}{\varepsilon_s} + \frac{1 - f_S}{\varepsilon_j}\right]^{-1}\end{aligned} \tag{2.24}$$

It can be seen that both zeroth and second-order EMT expressions are polarization dependent, which means that the 1D ARS has a property of structural (built-in) form birefringence.

For multilayer 1D ARS after introducing the effective refractive index and reducing a single layer to an equivalent homogeneous one, we may use the same recurrent relations as for the usual AR dielectric layers. A m-level film may be produced by using 2^m photolithographic steps. It is interesting that multilevel structures do not have to be symmetric, i.e., only the fill factor of a given level is important, not the geometrical position of the filling itself. Structures with a fill factor continually changing from the substrate to the top may be approximately represented as a large enough number of infinitesimally thin discrete and homogeneous rectangular elements.

For two-dimensional AR structures, literature quotes a number of different approximations (e.g., [162]). Lalanne proposed expressions for effective index in the form of infinite series [181]. Raguin [169] and dos Santos [182] showed that there is a direct correspondence between the optical properties of two-dimensional subwavelength ARS and homogeneous materials.

In practical calculations of 2D ARS (both single- and multilevel) and of 1D near cutoff (comparable Λ and λ), one may use the rigorous coupled-wave analysis (RCWA) [180]. This theory was used for the case of rectangular AR structures by Moharam and Gaylord [174, 183, 184]. The electromagnetic field for a planar structure is first expanded over harmonics and together with the Bloch periodicity condition replaced in the wave equation written over TE or TM polarized waves. In this way, an infinite set of coupled-wave equations is obtained [174]. By transforming this system and applying the boundary conditions for tangential components of electric and magnetic field, an infinite set of linear equations is obtained. Finally this set is truncated, and an arbitrary accuracy is obtained by retaining a sufficient number of equations. This is used to calculate the amplitude of certain orders of electromagnetic waves, and based upon this the diffraction efficiency of the structure is determined.

In the case of surface relief with an arbitrary shape, the relief is divided into a large number of homogeneous thin layers parallel with the surface. The above method is then applied to each of the layer and the boundary conditions are applied to the dividing surface between each pair of the planar layers. For a rectangular structure, this is reduced to a single layer.

The function of an ARS can be described by considering its profile, as shown in Fig. 2.23.

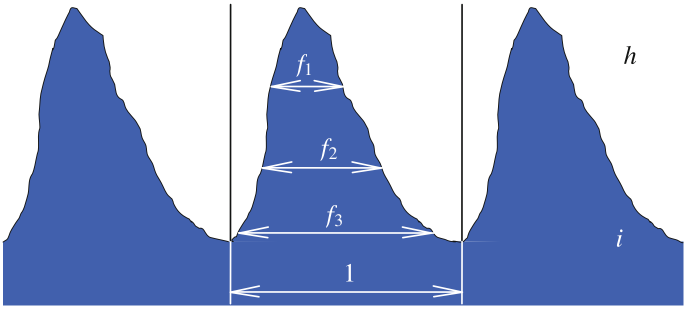


Fig. 2.23 Applying effective medium approach to diffractive antireflection structures. Here f_i denotes the filling factor at different heights, the largest value being 1 and the lowest 0; $f_1 < f_2 < f_3$

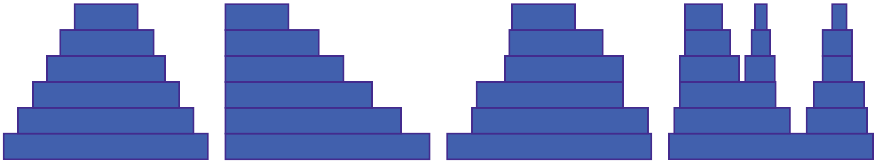


Fig. 2.24 Equivalent ARS structures with different profiles but with identical dependence of filling factor on height

It can be seen that the ratio of the amount of the ARS material (denoted by i) and the surrounding medium (denoted by h) changes with height, f being larger nearer the active surface (bottom). If one divides the ARS in infinitesimally thin horizontal sheets and applies the effective medium theory to each one of them, one sees that the relative dielectric permittivity increases gradually from about 1 on the top (since there $f = 0$) to the value of the active semiconductor permittivity at the bottom—i.e., the ARS shown in Fig. 2.23 effectively behaves as a graded refractive index structure, shown to be the most efficient planar AR structure [162].

Since the structures in question are subwavelength and observed as an effective medium with a given filling factor at a given height from the detector surface, it is then irrelevant which profile exactly it follows, as far as the filling factor does not change. In other words, the profiles shown in Fig. 2.24 are identical from the point of view of their properties as antireflective structures.

The range of acceptable incident angles for an ARS is very large and it can be designed for optimum operation for a wide range of operating wavelengths. Another important advantage is that the ARS can enable fabrication of any desired effective refractive index, including the values that do not exist in nature. Also, they do not pose the problem of material adhesion to the substrate and thermal expansion coefficient matching between the substrate and the ARS.

A relief ARS array may be one-dimensional (Fig. 2.25) or two-dimensional, Fig. 2.26. In both cases it may be fabricated with a continuous profile, but also as binary or multilevel. Most frequently mentioned in the literature are 1D ARS with a

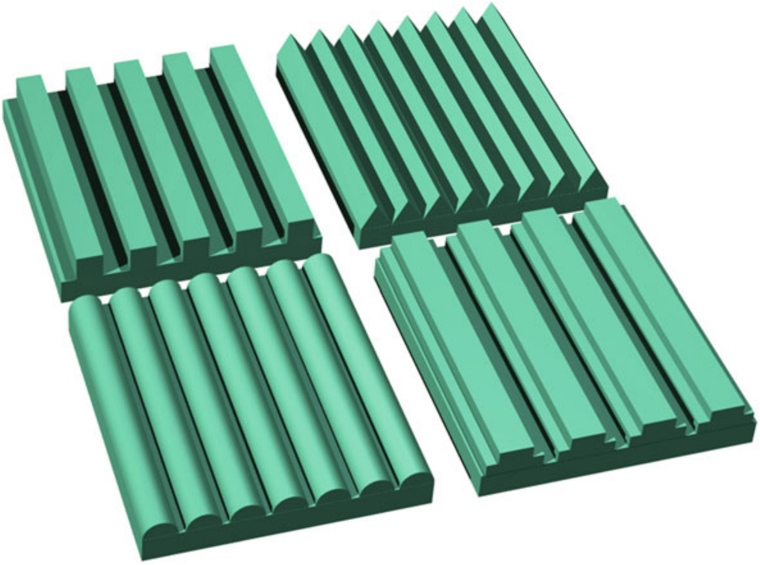


Fig. 2.25 Various 1D subwavelength ARS. *Top left* 1D rectangular structure; *top right* 1D sawtooth (triangular, V-groove) structure; *bottom left* cylindrical; *bottom right* 1D two-level, rectangular

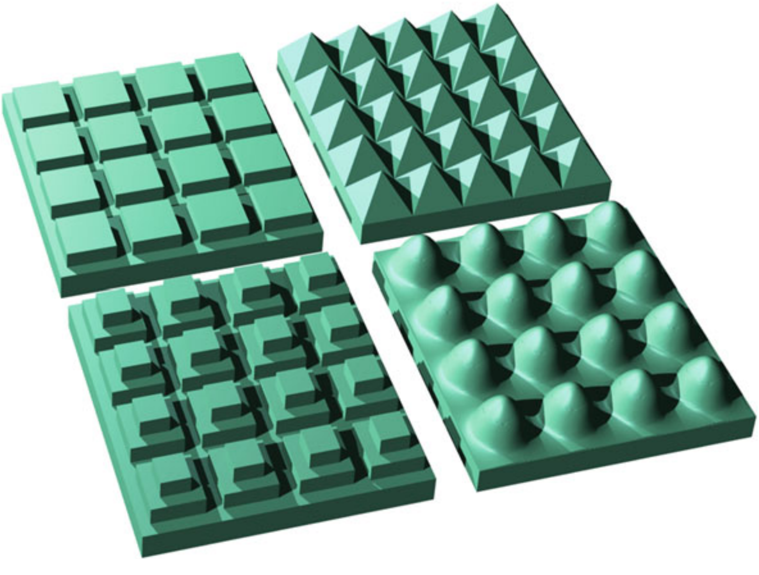


Fig. 2.26 2D subwavelength ARS. *Top left* parallelepiped profile; *top right* pyramidal shape; *bottom left* two-level parallelepiped (binary); *bottom right* 2D sinusoidal

rectangular relief profile [174], triangular ("toothed," "saw," "sawtooth," or "V-shaped") [185] and sinusoidal [119, 182]. Two-dimensional ("crossed") structures may be parallelepiped-shaped (single- or multilevel) [182], pyramidal (truncated or full) [186], sinusoidal [187], etc. Obviously, the structures shown in Figs. 2.25 and 2.26 are only a few of the basic versions and many other geometries are possible. For instance, one may use the "negatives" of each of the presented structures—where protrusions are shown, one can use instead holes of the same shape. Instead of pyramids shown in Fig. 2.26 top right, one could use pyramidal holes, etc.

Different technologies are available for the fabrication of micrometer or submicrometer-sized subwavelength ARS for middle- and long-wavelength infrared radiation. An ARS profile may be fabricated in resist utilizing photolithography, directly in the active area of the photodetector or in an amorphous buffer layer deposited for instance by sputtering. A variant is holographic lithography (being a standard method for the fabrication of Bragg gratings), used for submicron-pattern ARS. Direct-writing electron beam and focused ion beam lithography are also used.

To transfer a photolithographically formed pattern of an ARS into the active area or into a buffer layer techniques of wet and dry etching are used. Among the methods of wet etching, a frequently mentioned one is lift-off technology. In the processes of etching, sometimes additional buffer layers are used as an etch stop to accurately control etching depth. Much more often methods of dry etching are used, especially reactive-ion etching (RIE) and chemically assisted ion etching. Electron-beam lithography, ion milling and ion implantation are also used.

Numerous methods for nanofabrication were applied to fabricate patterned AR 3D profiles. Bataglia et al. used self-assembled nanospheres to perform nanosphere lithography by reactive-ion etching and obtain patterns of nanocavities on glass [188]. Spinelli et al. [189] produced "black" silicon by depositing silicon nanopillars 250 nm in diameter, 150 nm high on the silicon substrate. The nanopillars were ordered in a square 2D lattice with a 450 nm pitch. They acted as subwavelength resonant Mie scatterers, thus producing large scattering cross-sections with large forward scattering. The obtained reflection was 1.3 % in the visible. Like other ARS, the structure was omnidirectional.

One of the possible approaches for the photodetectors in (8–14) μm range (mercury cadmium telluride and similar zincblende crystals) is to fabricate a 2D array of openings by anisotropic etching through windows in photolithographic mask (along crystallographic axes). The method is somewhat similar to the one used to characterize etch-pit dislocations. In dependence on the crystallographic orientation and the etching solution used, the shape of the openings may be tetrahedral (orientation (111)Cd), trapezoidal (100)Cd, or pyramidal $(1\bar{1}\bar{1})$Cd. If preferential etching is performed without the aid of a mask, stochastically distributed tetrahedrons are obtained.

High-resolution (submicrometer) photomasks may be used to fabricate microopenings by isotropic etching of HgCdTe using Br-methanol solution. In this way calottes and hemispheres are obtained that may be separated or overlapping.

Diffractive ARS offer important advantages over interference AR coatings. They can enable fabrication of any desired effective refractive index. They do not pose the problem of material adhesion to the substrate and thermal expansion coefficient matching between the substrate and the ARS. Contrary to homogeneous dielectric films, diffractive ARS can be used for a much larger range of incident angles, exceeding 50°.

2.6.4 Bionic Approach: Moth-Eye Structures

Biomimetics, bionics, or bio-inspired design is a very important direction of fabrication of highly efficient engineering structures generally. Particularly, numerous works have been dedicated to biomimetics in photonics [190, 191]. It has been observed that some insects developed organic formations that act as very efficient antireflection structures. Some species of moths have their compound eyes covered by ordered arrays of submicrometer nipples. These structures, also variously called protuberances or nodules, are rounded at the top and arranged in a hexagonal lattice with a periodicity of about 250 nm. It has been found that they act as very efficient antireflection structures in a way similar to the previously mentioned antireflection diffractive surface relief with gradient effective refractive index profile. Very similar structures have been observed on the transparent wings of some hawkmoths, where they are responsible for their visual appearance of blackness, like a kind of natural "stealth" design. First considerations of such extraordinary properties of the eyes of moths came relatively early [192].

Figure 2.27 shows the appearance of a moth-eye surface profile. The structure exhibits strong antireflection properties for all polarization and is omnidirectional. Its bandwidth is extremely wide and covers the range from UV to the terahertz region.

Some literature sources go so far that according to them all subwavelength antireflection reliefs with a gradient of effective refractive index are called moth-eye structures (e.g., [193]), including 2D pyramidal and nanopillar arrays.

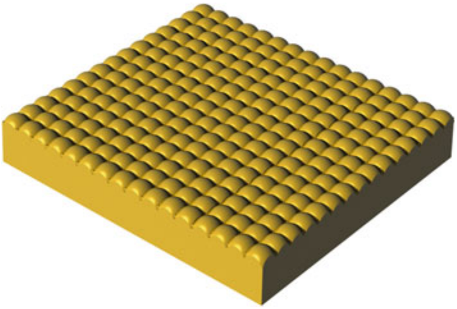


Fig. 2.27 Illustration of an artificial antireflection structure based on the moth-eye principle

Decades passed before it was technologically possible to mimic such nanostructures and fabricate artificial bionic moth eye ARS [191, 194]. Technologically, moth eyes can be produced at a correct scale using lithography, crossing three gratings at 120°. Another approach to fabrication is electron-beam etching.

Harker et al. produced artificial moth eye structures in silicon, germanium, and diamond [193]. Huang et al. fabricated bio-inspired silicon nanotips as antireflection structures [194].

One should mention here the works of Wang who implemented a topology optimization procedure by genetic algorithm to developed a complex optimized surface relief (not a moth-eye structure), as a result claiming efficiencies up to three times over the Yablonovitch limit [195].

2.6.5 Random Surface Corrugations

Besides the regular (patterned) profiles, there are a number of random or "stochastic" antireflective diffractive structures, i.e., structures utilizing arbitrarily distributed elements (whose shapes themselves may be regular or not). Usually their behavior is similar to those with pyramidal or sinusoidal relief, because their effective density gradually increases from the ambient toward the substrate. These stochastic reliefs may consist of randomly distributed smaller or larger geometrical shapes obtained by substrate etching, e.g., tetrahedrons, rounded "pillars" [196], hemispheres, etc. or the geometrical shapes themselves may be fully random (Fig. 2.28).

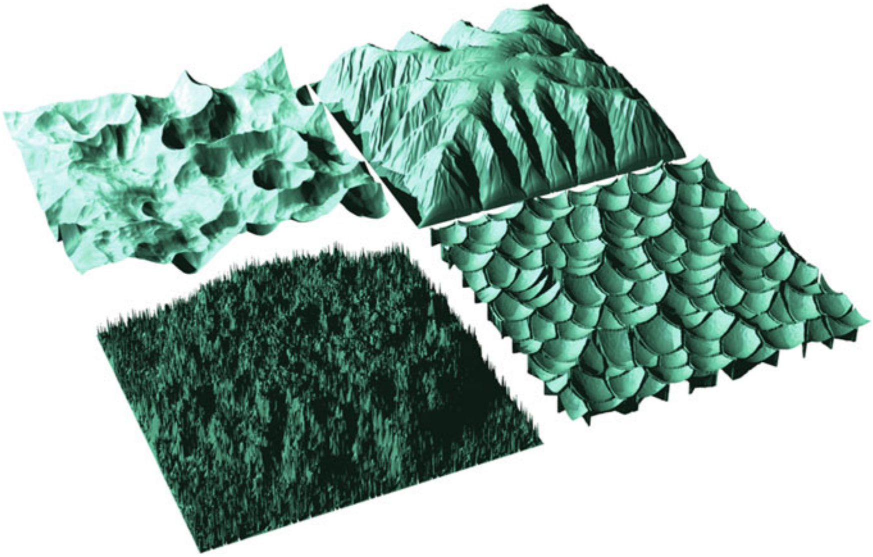


Fig. 2.28 Some stochastic surface reliefs for ARS

The question which kind of antireflection surface relief is the most efficient, regular or random, is still open to debate [188]. Some of the most efficient detectors in the solar cells field actually have random surface profiles—stochastically distributed pyramids made in transparent conductive oxide (zinc oxide) using chemical vapor deposition [197]. However, ordered and random profiles both effectively behave as gradient refractive index profile starting from the effective value nearly equal to that of the environment index and ending with that of the substrate. In this manner, the question reduces to the determination of an optimal gradient-index profile.

The scattering properties of surface profiles can be described by two main parameters, the haze and the angular distribution functions [198]. The haze is defined as the portion of the light that is diffusely (i.e., nonspecularly) scattered into the far field, while angular distribution function describes angularly resolved scattering. Rockstuhl et al. [199] made a connection between these parameters and the absorption properties in the active region and performed a numerical calculation to determine the optimum stochastic surface profile. The conclusion is that a random surface relief should scatter light as strongly as possible and that simultaneously the angular domain of that scattering should be as large as possible in order to ensure maximum transmission.

To create "stochastic" subwavelength antireflection relief anisotropic etching of single crystalline layer was used [200, 201]. Another method is reactive sputtering in plasma [196]. Some authors utilized chemical vapor deposition [197]. Random arrays of core-shell nanowires grown by vapor–liquid–solid method were reported in [202]. For HgCdTe, wet chemical etching and dry etching in plasma were used [203]. The most often used wet etchant is the solution of bromine in methanol. Other solutions of bromine in alcohols (butanol, ethylene glycol) or acids are also utilized. These methods are preferred since they do the least damages to the HgCdTe crystal lattice. Various dry etching solutions are applied as well, using plasma based on methane, hydrogen, argon, and nitrogen, sometimes with added halogens like fluorine or chlorine (convenient for stochastic surface profiles).

A problem with random profiles is their repeatability from sample to sample, since it depends on a number of process parameters and their very nature is stochastic. Some solutions that could be applied to overcome this is to use advanced replication technologies like, e.g., nanoimprint to replicate the optimized random profile. Another is to apply top-down planar fabrication to pattern quasiperiodic structures that behave as pseudorandom, although they are actually deterministic.

2.6.6 Subsurface Scatterers

It is possible to increase the optical path through a photodetector by incorporating different structures within the active region itself so that they act as scatterers of light. In this way, a beam that would else leave the detector remains within the absorptive domain. Structurally, one could use inclusions, fillers, pores, voids, etc.

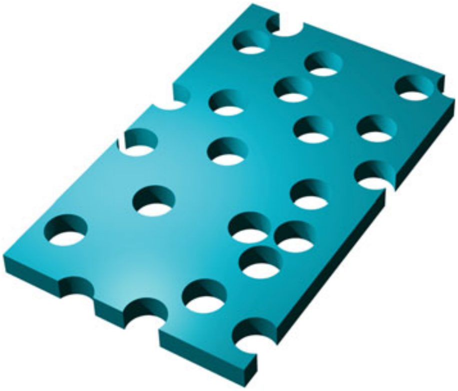


Fig. 2.29 Disordered 2D array of nanoholes for light trapping in thin absorptive media

with subwavelength dimensions. After entering the active region, light beams are diffracted on scatterers which increases their overall optical path and effectively traps them. This mechanism is actually similar to the scattering of the internal beams within the detector from the surface antireflection reliefs. Like in the case of other structural arrays for detector enhancement, the subsurface scatterers may be ordered/periodic, quasiperiodic (with a benefit that one has a deterministic and repeatable structure that behaves as a quasi-random one), aperiodic or fully random.

A method of the use of scatterers incorporated into the absorptive region was described by Vynck et al. who produced a random array of nanoholes in absorptive film [204]. Multiple scattering here is two-dimensional and leads to an appearance of disordered quasi-guided modes. Their structure is schematically presented in Fig. 2.29.

2.7 Metamaterials as Antireflection Structures

Compositing of materials at subwavelength level brought to a novel class of structures known as electromagnetic metamaterials [205, 206]. They could be defined as artificial structures whose effective electromagnetic properties exceed those found in nature. By tailoring their geometry and material properties, one is able to directly influence their frequency and spatial dispersion. In this manner, it is possible to obtain effective relative dielectric permittivity, relative magnetic permeability, and refractive index in a vastly expanded range compared to naturally occurring materials. Group velocities in metamaterials can be designed to be much larger than in free space (the superluminal group velocity—the so-called “fast light,” corresponding to very low values of refractive index, including those near zero), much smaller (“slow light“—very high effective refractive index), and even negative (“left-handed light“—effective refractive index below zero) [207].

Structurally, metamaterials are usually metal-dielectric composites ordered in 1D, 2D, or 3D arrays, where a certain metallic pattern (denoted as electromagnetic

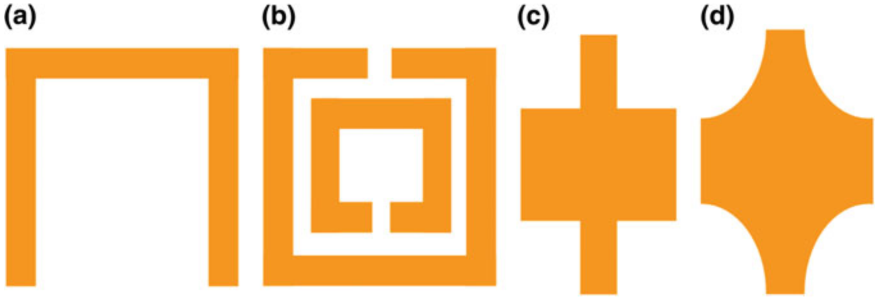


Fig. 2.30 Some metamaterial building blocks ("particles"). **a** Single split ring, **b** double split ring, **c** square fishnet, **d** elliptical fishnet

"atom", "molecule" or "particle" because of the analogy with natural particles and their electromagnetic response) is designed to furnish a desired response and then repeated to ensure the designed effective properties, Fig. 2.30. This is why such artificial "atoms" (which obviously comprise a very large number of natural atoms) must be much smaller than the operating wavelength in free space. Engineering of group velocity is obtained by adjusting their capacitive or inductive behavior, i.e., their resonant/antiresonant response to electromagnetic radiation.

The most often met class of metamaterials are those with negative effective refractive index. These are structures in which effective permittivity and permeability are negative in the same frequency range (the "double-negative" materials). Envisioned as a curiosity in 1967 by Veselago [208], practically introduced by Pendry [209] and experimentally fabricated by Shelby et al. [210], they have a plethora of unusual properties. The direction of the Poynting vector in negative index metamaterials is opposite to that of the wavevector, i.e., electric, magnetic, and wave vectors form a left-oriented triplet. Hence they are often called "left-handed materials".

It has been suggested that negative index metamaterials can be used for the fabrication of near-field lenses that exceed the diffraction limit and operate in subwavelength range (perfect lenses and superlenses) [211]. A whole new subfield of electromagnetic optics has been derived from the paradigm of metamaterials, the transformation optics [212, 213]. One being able to tailor optical response at will, it is in principle possible to perform a mapping of the optical space from one state to another, thus obtaining a desired electromagnetic behavior. Some of the proposed applications are invisibility cloaks [214, 215], superconcentrators [216], hyperlenses [217], etc.

Since the properties of metamaterials and metamaterial-containing structures can be engineered almost at will, an obvious question that occurs is if it is possible to use them as superior impedance matching structures. The problem reduces to the use of metamaterials as impedance matching structures between the surrounding medium and the material of the detector-active region.

2.7.1 Metamaterial-Containing Antireflection Multilayers

The simplest way to use metamaterials as antireflection structures is to make a 1D multilayer in which isotropic layers of negative index metamaterial and lossless dielectric are alternating. Such structures were extensively researched [218–222]. Metamaterial-containing multilayers are characterized by many peculiar properties with both theoretical and practical interest. Using such structures it is possible to fabricate resonant cavities with subwavelength dimensions [223].

Compared to conventional Bragg-type quarterwave-stack interference multilayers that exhibit a single stop-band (1D photonic band gap) where propagation of electromagnetic waves is forbidden, metamaterial-containing multilayers have three additional kinds of stop bands besides such Bragg stop-band. One of the novel stop bands appearing in alternating metamaterial-dielectric stacks is related to zero-value of refractive index averaged over volume of the multilayer and is denoted as zero-index band gap [220]. Another one is in a similar manner related with zero permittivity and the third one with zero permeability bandgap.

The new band gaps appearing in metamaterial-containing stacks only are insensitive to change of scale. They are also insensitive to disorder, and remain even in quasiperiodic structures [224, 225]. Even more important, it has been shown that metamaterial-containing multilayers do not change their properties with a variation of incident angles, i.e., they exhibit omnidirectional behavior. Most of such metamaterial properties are consequences of phase compensation [226]. Namely, the metamaterial slab fully cancels the phase of a wave arriving from the dielectric part, so that each pair of strata can be made completely transparent. Due to the same effect, higher and wider bandgaps appear in such structures than in conventional quarterwave stacks.

It has been suggested that multilayer structures comprising of alternating dielectric and negative index metamaterial could be used as a novel kind of antireflection coatings [227]. Compared with conventional Bragg-type all-dielectric interference films, these structures are much less sensitive to the incident angle, which could improve the characteristics of antireflective coatings [227].

A problem with such use of metamaterials are large absorption losses in the range of negative refractive index and relatively narrow bandwidths due to their resonant nature. Technologically, the fabrication of metamaterial structures is a nontrivial task. For the visible and infrared range it is typically done by top-down approach (planar technologies and dry etching), but the necessary resolutions are deeply subwavelength. A multilayer metamaterial based on the so-called double-fishnet structure has been reported by Valentine et al. [228].

2.7.2 Metamaterial Superabsorbers

Plasmonic metamaterials can be utilized to match the impedance of environment (air) to the substrate (semiconductor material) in the optical range (or actually any other range) directly by optimizing their "particles" design. Thus one can obtain "perfect black" surfaces denoted as superabsorbers [229, 230]. Such nanostructured surfaces are able to absorb light in a wide bandwidth regardless of its polarization and for literally any incident angle.

Basically, all superabsorbers proposed until now were metal-dielectric–metal sandwiches where the top metal layer of the sandwich is nanostructured to obtain a desired wave impedance.

Figure 2.31 shows two typical structures for superabsorbers. The bottom one represents an ordered array of plasmonic nanoparticles on dielectric. The shape of the nanoparticle array may be rectangular, circular, etc. [230] The top structure represent a 2D array of nanoholes in the top metal sheet [231]. In both cases dielectric is located on a continuous metal sheet.

Another structure represents 1D trapezoidal or crossed 2D arrays (Fig. 2.32). The basic form of this super absorber is again metal–dielectric–metal [232].

As far as photodetectors are concerned, superabsorber structures were described within the context of photovoltaic solar cell enhancement [232]. However, other wavelength ranges can be covered by changing the dimensions of the elements of the superabsorber.

Technologically, the metal-dielectric-metal sheet with satisfactory properties can be fabricated utilizing conventional radiofrequent sputtering. Nanostructuring of the

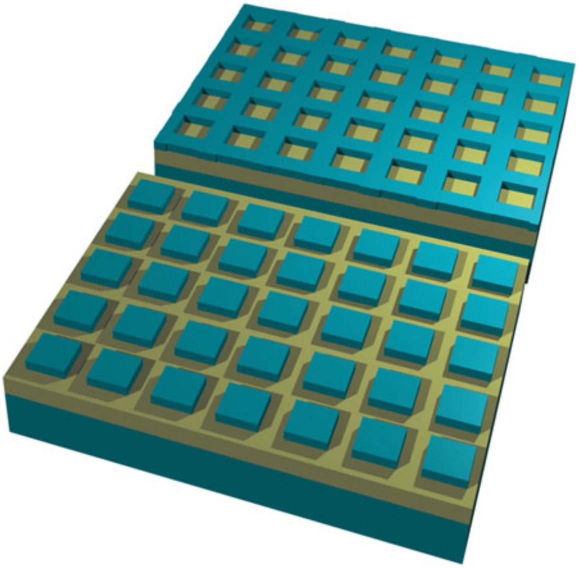


Fig. 2.31 Plasmonic metamaterial-based super absorbers. *Bottom* Metal nanoparticles on dielectric; *top* Nanohole array in metal layer on dielectric. *Bottom* layer is metal

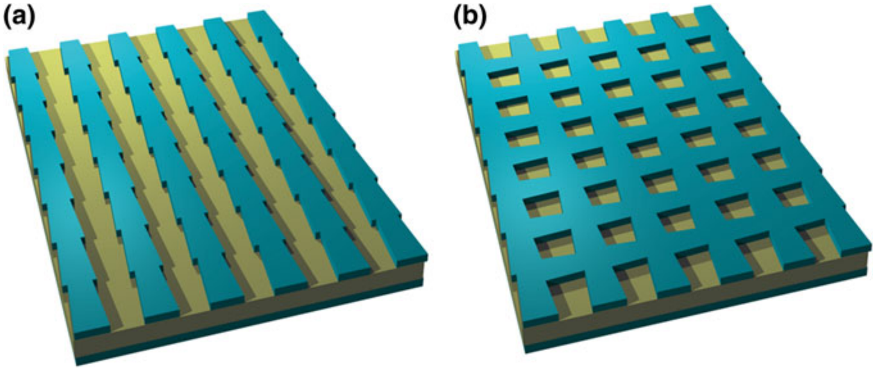


Fig. 2.32 Plasmonic super absorber based on trapezoidal plasmonic elements. **a** 1D array; **b** 2D array. Metal layers are on the *top* and the *bottom*, divided by dielectric

top surface can be done utilizing photolithography and electron-beam or focused ion beam methods [230]. In addition to that, structures of the type shown in Fig. 2.31 bottom may be fabricated utilizing functionalization by plasmonic nanoparticles [233]. In that case disordered nanoparticle arrays are obtained, but the effective properties of the structures remain similar to the ordered ones for similar values of the filling factor f.

Alici et al. described a metamaterial absorber based on a 2D array of single split ring resonators [234, 235]. The structure was designed for the infrared range. Sun et al. designed a metamaterial absorber with wide operating range by using several layers, each with split ring resonators with different dimensions [236].

2.8 Light Trapping by Optical Path Increase

A possible way to increase the absorption probability while at the same time keeping the detector optical thickness αd constant is to ensure a longer path of light beam through the detector. The idea is to trap the light that has already entered the active region and keep it as long as possible. The most general way to implement such light trapping is to influence the path of the beam, i.e., modify the optical space in some manner so that the radiation continues to propagate within the detector. This could be done in several ways. The beam could be returned into the detector from its surface, immediately before leaving the structure. It could be made to scatter and to repeatedly change its path within the structure. It could be also concentrated within the structure itself, by making some kind of internal hotspots through the influence to the local density of the optical space.

The simplest and the most obvious approach is to utilize some kind of high-reflection structure to change the path of the beam and return it back to the active

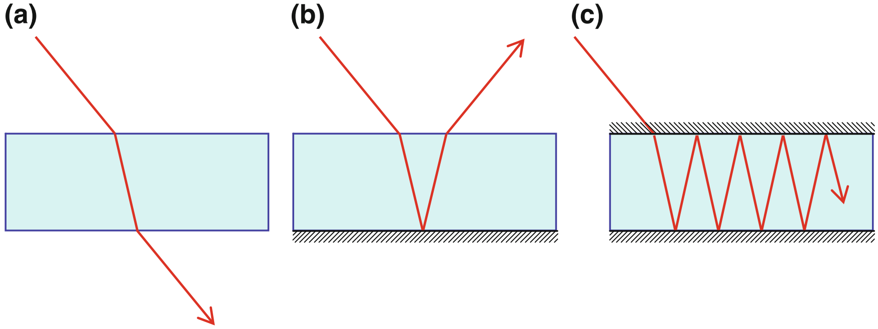


Fig. 2.33 Idealized representation of optical path increase in photodetector using high-reflection surfaces. **a** No high-reflection surfaces are used: light refracts and leaves the detector after a single pass; **b** backside reflector used: light passes the active region twice; **c** cavity enhancement is used: light continues to reflect multiply

region from the exit path. Figure 2.33 shows an idealized representation of optical path through a thin semiconductor detector without light trapping and with two different schemes utilizing high-reflection surfaces.

Figure 2.33a shows a photodetector before any light trapping scheme is implemented. It is assumed that an antireflection coating has been applied to the detector surfaces, ensuring almost complete transmission of the incident beam into the active region of the photodetector. In a real structure, a part of the radiation would be reflected both on the front and on the back surface due to the refractive index mismatch between the surrounding medium (low index) and semiconductor material (high index), and no AR coating would ensure 100 % transmission. Obviously, a part of the beam intensity would be lost due to the absorption within the structure, the effect we desire to maximize. The beam refracts at the front and at the back surface and is immediately lost after a single pass.

A different situation is met if a high-reflection layer is deposited on the detector backside, as shown in Fig. 2.33b. The signal is completely reflected from it and passes a twice longer path through the active region before leaving the structure. This solution is very often implemented and represents almost a rule for practically all conventional thin photodetectors. In real situations, there are again multiple reflections due to refractive index mismatch. A calculation that takes into account these reflections has been presented in [11].

The efficiency can be further improved if there are high-reflection surface both on the front and the backside of the detector, Fig. 2.33c. In other words, the detector is placed into a cavity with highly reflective walls. Light will enter into such structure resonantly or through some kind of light pipe, as described later in this text. Upon entering, the beam will reflect multiply and will be kept within the active region much longer. Large enhancements ensue as a result. This is the basis of the cavity enhancement of detectors. The cavity itself may support one or a larger number of modes and may be implemented in different ways. Examples of cavity enhancement include resonant cavity enhancement (RCE) and photonic crystal enhancement (PCE) [237].

The next section analyzes different high-reflection structures convenient for cavity enhancement. It also considers strategies and conditions of their use.

2.9 High-Reflection Structures

2.9.1 Introduction

A reflective surface can be specular (mirror-like, glossy) or diffuse, Fig. 2.34. The retroreflective surfaces return the beam in the direction the incident light came from. Regardless of the type of reflectivity, most of the radiation is returned from the illuminated surface toward the radiation source. This radiation can be practically unchanged or can be shifted in frequency spectrum, either toward higher or toward lower frequencies (“hot” and “cold” mirrors).

The basic parameter of a reflective surface is its reflection coefficient, *R*, describing the percent of the incident radiation returned from the reflective surface. The largest values of *R* are offered by multilayer dielectric films. Reflection coefficient is usually dependent on light wavelength, and there is a spectral range of reflection maximum for a given type of material. Typically, metals have a very wide spectral range of high *R*, while for dielectric mirrors this range depends on the properties of a particular structure and is typically much narrower. On the other hand, dielectric mirrors offer much higher values of the reflection coefficient.

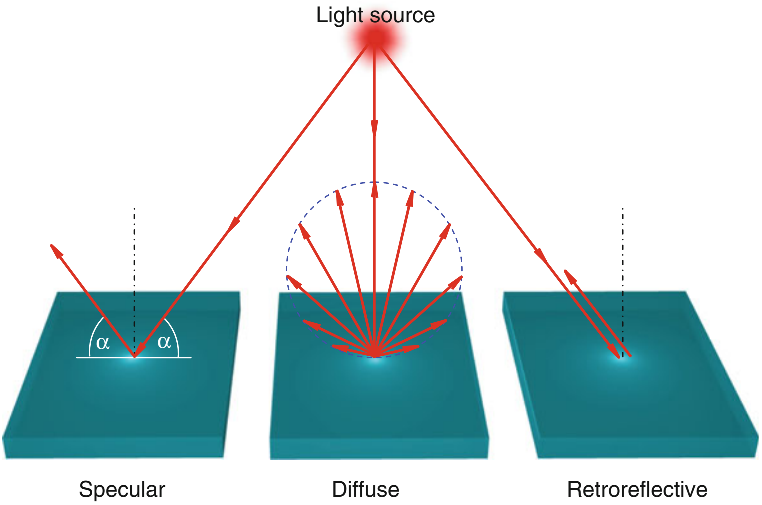


Fig. 2.34 Different types of reflectivity: specular (mirror-like), diffuse, and retroreflective

Another issue of interest is the angular dependence of reflection coefficient. For some structures, this dependence is very strong and large values of R are obtained only for near-normal incidence, thus furnishing surfaces with very high directivity of reflected light. The other pole are the surfaces offering near-omnidirectional or even fully omnidirectional reflectivity.

A surface can have significant diffuse reflectance, dependent on the scattering properties of the material. There is a range of possible intermediate situations between a fully specular and fully diffused optical reflector surface.

In this text, we handle different types of high-reflectance structures that could be of use for photodetector enhancement through optical path increase, starting from the simplest and well-known ones like metal mirrors and continuing to more complex ones like interference mirrors and ending with structures like photonic crystals. The function of dielectric mirrors and photonic crystals is based on light wave interaction effects (interference and diffraction). Both utilize nanoscale-patterned structures and the effects of multiple reflection on structures with periodic changes of refractive index. They are part of a larger field of the so-called "designer" nanostructured materials that represent the current limits of materials engineering.

2.9.2 Specular Reflectivity

The specular type of reflectivity means that scattering on a given optical surface is low or zero. The angle of incidence with respect to the target surface is equal to the angle of reflection (Fig. 2.34). The incident beam and the reflected beam lie in the plane defined by the incident beam and the normal to the surface. In other words, the surface is mirror-like. Only the macroscopic curvature of the surface may influence the shape of a reflected optical image.

Specularly reflective surface can be used in image forming. It should be mentioned that this is not of prime importance for photodetectors, since no image is formed in a single photodetector element, and the only important factor is the value of reflection coefficient for different wavelengths and incidence angles.

The conventional solution to achieve specular reflectance is to use flat metal surfaces. Other solutions are interference-based multilayer dielectric reflectors (Bragg mirrors) and, as their generalization, photonic bandgap structures (photonic crystals) of all-dielectric and metal-dielectric type, etc. Nanoscale interferometric and diffractive structures offer extremely large values of reflection coefficient (in excess of 99.99 %).

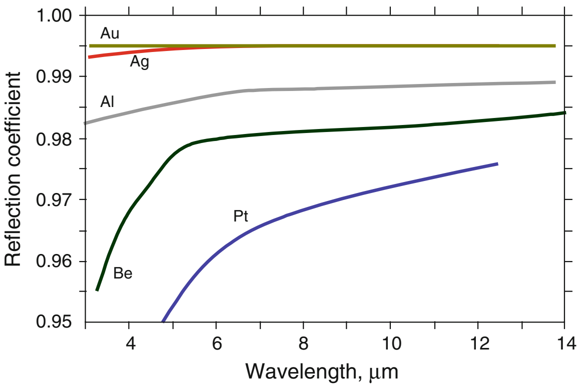


Fig. 2.35 Spectral dependence of reflection coefficient of some metals used for reflective layers in IR detectors

2.9.3 Metal Mirrors

The spectral range of thin metal mirrors is typically very wide. Also, they are relatively insensitive to the changes of the incident angle. Because of their low cost and relatively high-reflection coefficient they are often used for the deposition to the backside of the detector. Since thin metal mirrors quickly become covered by an oxide layer spoiling their reflection, in some applications additional dielectric protective coatings are deposited on them, for instance magnesium fluoride or silicon monoxide. This is usually not necessary when they are used for the improvement of photodetector quantum efficiency.

The most often used materials for the infrared range are aluminum and gold [5]. There is a wealth of literature data on these materials in the IR range. As an example, Fig. 2.35 shows spectral characteristics of the reflection coefficient of some polished metals in the (3–14) μm range, according to [1, 238–240]. All dependences are valid for normal incidence and at room temperature.

One of the important parameters for the use of metal reflectors in MOEMS (microoptoelectromechanical systems) is the penetration depth (skin depth) of electromagnetic radiation at a given wavelength. In the (3–14) μm range, its value is typically between below 10 nm (aluminum) and 20 nm (nickel) [240].

2.9.4 Dielectric Interference Mirrors

Interference mirrors are dielectric thin film coatings where low- and high-refractive index layers alternate. The optical thickness of each of the layers is equal to quarter-wavelength ($\lambda/4n$). They are denoted as distributed Bragg mirrors or distributed Bragg reflectors (DBR), sometimes simply as Bragg mirrors. Other names include quarter-wave mirrors (QWM), quarterwave stacks (QWS) and highly reflective (HR) layers.

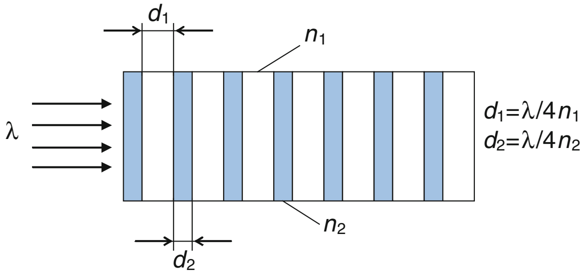


Fig. 2.36 Schematic presentation of a high-reflection quarterwave stack

Bragg mirrors are routinely fabricated by alternatively depositing quarter-wavelength stacks with high (n_h) and low (n_l) refractive index (approximately lossless dielectrics are assumed). Their principle of operation is similar to that of multilayer antireflection films, but the incident beam arrives at the layer with high-refractive index, and in this case the interference on the quarter-wavelength stacks is constructive.

The theory of dielectric mirrors belongs to the general theory of optical interference coatings [152, 163, 180]. The simplest type of specular multilayer reflectors are high-reflection (HR) dielectric quarterwave stacks. These are structures consisting of a number of micrometer- to nanometer-thick dielectric layers. The optical thickness of each separate layer of such a structure (the ratio of real thickness and refractive index) is equal to a quarter of the operating wavelength (thus the name “quarterwave stacks”) and the value of refractive index is periodically alternating between high (n_2) and low (n_1) values (See Fig. 2.36).

After a light beam enters the stack, each interface between two adjacent layers will reflect a part of the incident radiation. All the emerging beams will be in phase and thus their intensities will add. In that way constructive interference will greatly enhance the reflection of the whole stack.

An approach that can be generally applied to various multilayer structures is the well-known transfer matrix technique [87, 180]. It can be used regardless of the geometry of the layers or the intended application of the multilayer. Besides being usable for the calculation of quarterwave Bragg mirrors, it can be used without modification to accurately determine the properties of antireflection coatings, step-down structures, various quasiperiodic, aperiodic and random structures, but also 2D and 3D photonic crystals as well [241]

In comparison to an arbitrary multilayer coating, a quarterwave 1D dielectric stack has additional symmetry which allows one to determine its transfer matrix analytically, thus significantly simplifying the procedure of calculation of its properties and optimization of the final structures.

To apply the transfer matrix method to a quarterwave stack the electric field in each separate homogeneous layer of the stratified Bragg mirror structure is written as a superposition of the incident ($E_j^{(i,l)}$) and reflected ($E_j^{(r,l)}$) plane wave, where j denotes the number of the pair (period) of the quarterwave mirror, $l = 1, 2$ is the number of the stratum within the given pair, r denotes the reflected wave, and

i incident. Assuming that a plane wave has Bloch (Floquet) periodicity and applying the condition of continuity of electric and magnetic fields and their first derivatives on the interfaces between strata, the connection between the j-th and (j–1)-th layer is [178–180]

$$\begin{bmatrix} E_j^{(i,l)} \\ E_j^{(r,l)} \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} E_{j-1}^{(i,l)} \\ E_{j-1}^{(r,l)} \end{bmatrix} = [P] \begin{bmatrix} E_{j-1}^{(i,l)} \\ E_{j-1}^{(r,l)} \end{bmatrix}, \tag{2.25}$$

where $[P]$ denotes the transfer matrix with the elements

$$\left.\begin{aligned} P_{\substack{11\\12}} &= e^{e\pm ik_1 d_1}\left[\cos k_2 d_2 \pm \frac{i}{2}\left(\frac{n_2}{n_1}+\frac{n_1}{n_2}\right)\sin k_2 d_2\right] \\ P_{\substack{12\\21}} &= e^{\pm ik_1 d_1}\left[\pm\frac{i}{2}\left(\frac{n_2}{n_1}-\frac{n_1}{n_2}\right)\sin k_2 d_2\right] \end{aligned}\right\} k_1 = \frac{2\pi}{\lambda} n_l, \quad l = 1,2\,. \tag{2.26}$$

For an M layer structure the Chebyshev equality can be applied [87] to obtain

$$\begin{bmatrix} E_0^{(i,l)} \\ E_0^{(r,l)} \end{bmatrix} = \begin{bmatrix} P_{22}U_{M-1} - U_{M-2} & -P_{12}U_{M-1} \\ -P_{21}U_{M-1} & P_{11}U_{M-1} - U_{M-2} \end{bmatrix} \begin{bmatrix} E_M^{(i,l)} \\ E_M^{(r,l)} \end{bmatrix}, \tag{2.27}$$

where

$$U_M \frac{\sin[(M+1)\kappa d]}{\sin(\kappa d)} \tag{2.28}$$

is the Chebyshev polynomial of the second kind, and κ is the Bloch wavenumber

$$\kappa = \frac{1}{d}\arccos\left(\frac{P_{11}+P_{12}}{2}\right). \tag{2.29}$$

The Fresnel reflectance of the structure is [180]

$$r_M = \frac{P_{21}U_{M-1}}{P_{11}U_{M-1} - U_{M-2}}. \tag{2.30}$$

In the case when the refractive index of the incident medium n_a is not equal to n_1, we utilize [242]

$$r = \frac{\frac{|n_a - n_1|}{n_a + n_1} + r_M}{1 + \frac{|n_a - n_1|}{n_a + n_1}}. \tag{2.31}$$

The total reflection coefficient is calculated as the square of the absolute value of the Fresnel reflectance (2.31).

The reflection range of a quarterwave mirror is determined as

$$\frac{\Delta\omega}{\omega} = \frac{4}{\pi}\arcsin\frac{|n_2 - n_1|}{n_2 + n_1}. \tag{2.32}$$

The procedure of transfer matrix calculation of dielectric mirrors is described in detail in, e.g., [243]. An approach similar to the above is used for 2D and 3D structures, but the obtained expressions are much more involved and must be solved numerically. In [244] a detailed approach is described and the diagrams of photonic bandgap structures for 2D and 3D photonic crystals are given. An alternative method is presented in [241].

As an example of the application of transfer matrix method to the calculation of spectral reflection of quarterwave stacks, Fig. 2.37 shows the spectral reflection of a silicon-silica Bragg mirror for different numbers of layer pairs under normal incidence of light. A stop band is readily recognizable already for a small number of layer pairs. It is important that variations of the layer thickness and the materials used.

Reflectivity is greatly increased in a dielectric stack with a larger number of layers. A typical spectral dependence of a dielectric mirror is shown in Fig. 2.37. The wavelength range of maximum reflectivity is called the stop band (sometimes also called one-dimensional photonic band gap). Increasing the ratio between the low and high-refractive index parts extends the width of the stop band.

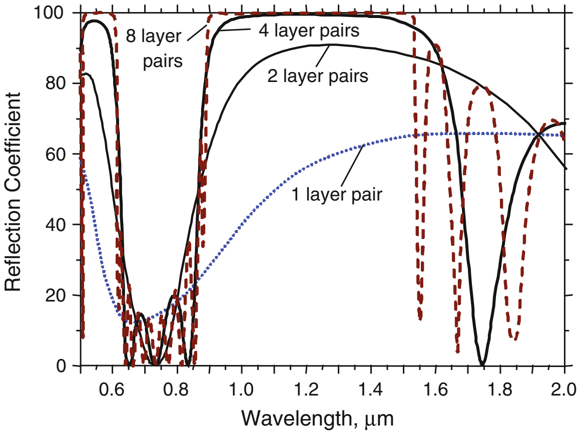


Fig. 2.37 Spectral reflectance of a HR dielectric stack for different numbers of layer pairs under normal incidence of light. Materials are Si ($\varepsilon = 11.958$, thickness 174 nm) and SiO_2 ($\varepsilon = 2.1316$, thickness 75 nm)

For some applications it is of interest to extend the high-reflectance zone (the stop band) and cover a larger portion of spectrum than in the case of quarter-wave stacks. There are several strategies to that purpose that deviate from the quarterwave stack concept. One of them is to stagger the thickness of each successive layer in some kind of progression (e.g., harmonic, arithmetic, geometric, etc.) [152]. Theoretically this could be also done by step-grading the refractive index from layer pair to layer pair, but the problem is the choice of materials. Chirped mirrors with Bragg wavelength varying between layer pairs in, e.g., linear fashion offer extended high-reflection range and at the same time enable control of the group delay [245, 246]. The stop bands in dielectric mirrors can be also extended by superimposing two or more quarterwave stacks designed for maximums in different wavelength ranges and to choose their ranges to be continuous. The proposed solutions include the use of dielectric stacks with gradient change of layer thickness and the use of layers with stochastically distributed thickness.

2.9.5 Omnidirectional Reflectors

The concept of omnidirectional mirrors, also denoted as perfect mirrors, was introduced in 1998 [247, 248]. These are one-dimensional, all-dielectric mirrors (1D photonic crystals) that keep their reflective properties even for large oblique incident angles in a wide and fully tailorable range of frequencies.

Basically, omnidirectional mirrors are dielectric quarterwave stacks in which TIR and Bragg mirroring wavelength ranges overlap, i.e., an overlapping band gap regime exists that extends above the light cone.

For a normally incident beam the stop bands for transverse magnetic (TM) and transverse electric (TE) polarization is identical. For increasingly oblique incidence the TM gap decreases, and the TE gap increases.

The width of the omnidirectional frequency range can be approximated for quarter-wave stacks as [249]

$$\frac{\Delta\omega}{2c} = \frac{(d_1 + d_2)\cos\left(-\sqrt{\frac{K_1-2}{K_1+2}}\right)}{d_1 n_1 + d_2 n_2} - \frac{(d_1 + d_2)\cos\left(-\sqrt{\frac{K_2-2}{K_2+2}}\right)}{d_1\sqrt{n_1^2 - 1} + d_2\sqrt{n_2^2 - 1}} \tag{2.33}$$

where c is the speed of light in the ambient, d_1 and n_1 are the thickness and the refractive index of the first layer in the pair, while d_2 and n_2 are the corresponding values for the second pair.

Figure 2.38 shows the spectral dependence of reflection coefficient for various incident angles for an Si/SiO$_2$ omnidirectional stack (dielectric permittivity 11.958/ 2.1316, thickness 174 and 75 nm, respectively, 8 layer pairs). The stop band changes with incident angle, becoming narrower and shifting toward shorter wavelengths with increasing incident angle. However, it does exist for all

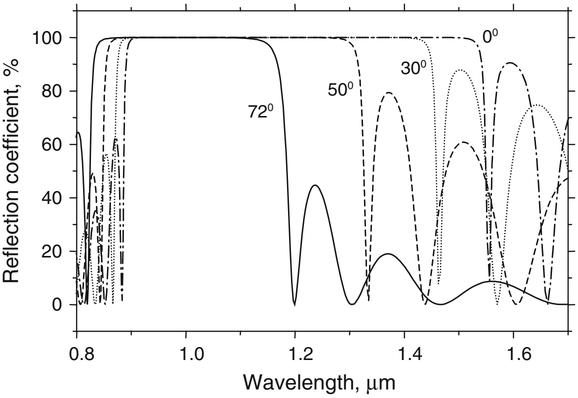


Fig. 2.38 Spectral reflectance of omnidirectional stack with 8 layer pairs for various incident angles, TM mode. Material pair is Si/SiO$_2$ (dielectric permittivity 11.958/2.1316), thickness 174 and 75 nm, respectively

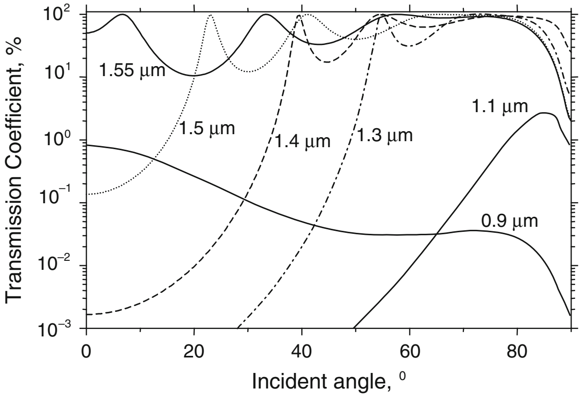


Fig. 2.39 Transmission coefficient of omnidirectional stacks with 8 layer pairs versus incident angle for different wavelengths, TM mode. Material pair is Si/SiO$_2$ (dielectric permittivity 11.958/2.1316), thickness 174 and 75 nm, respectively

considered incident angles, thus ensuring full functionality of the mirror for a wide range of oblique incidences.

Figure 2.39 shows the dependence of the transmission coefficient of omnidirectional 1D stacks on the incident angle for various wavelengths within the photonic band gap. The allowed angles decrease with the operating wavelength increase. For wavelengths outside the photonic band gap (1.55 µm) the transmission is high for all angles, remaining between 10 and 100 %. The lowest transmissions and highest reflections are obtained for the shortest operating wavelengths within the PBG.

The stop band (1D PBG) is fully tailorable to different wavelengths through changes of the materials used as high and low permittivity parts and through changes of the thickness of layers. Fink et al., who first described omnidirectional mirrors, performed experiments with polystyrene and tellurium and showed omnidirectional reflection in the LWIR wavelength range 10–15 μm [247]. Silicon and silica were used for the 3–5 μm MWIR range [250]. Numerous other material pairs are available for different wavelength ranges.

2.9.6 Metal-Dielectric Mirrors

Pure metal mirrors are omnidirectional reflectors, but the price to pay are their non-negligible absorption losses. On the other hand, all-dielectric multilayers exhibit very high reflectivity, but their reflectivity drops for oblique incident angles. Thus the idea occurs to combine both into a single structure.

A very old and obvious solution is to deposit a dielectric layer onto metal, thus reducing the absorptance and increasing reflectance. The transmittance remains very low or zero in such structures, but this fact is of no consequence for light trapping structures for photodetectors.

A generalization of this approach is to deposit a quarterwave dielectric mirror onto a metal reflector. In this way, very high reflectance is obtained for near-normal incidence. For more oblique incident angles the all-dielectric part will be transmissive, but in that case the metal part will reflect these beams, albeit with a lower reflectance. In this way the benefits of both types of the used structures are simultaneously employed. In [251] the authors mention the use of hybrid reflectors consisting of a Bragg-type dielectric mirror with an additional thin metal reflector deposited on its backside.

Another type of structures for metal-dielectric mirrors are metallodielectric multilayers. In this case alternating quarterwave or subwavelength stacks of metal and dielectric are deposited. Typical for such multilayers is a low reflection in visible, but large in infrared wavelength range. Thus they basically behave as low-pass optical filters. Such structures were denoted in literature as heat mirrors. First heat mirrors were fabricated as early as in 1950s [252]. The simplest heat mirrors consist of three layers only, dielectric-metal-dielectric or, alternatively, dielectric-transparent conductive oxide-dielectric [253]. Full multilayer metal-dielectric reflectors with binary but also ternary layers were also considered [254]. Because of their high reflectance in infrared, but also because of their plasmonic properties [255] metal-dielectric multilayer mirrors are of interest for cavity enhancement of infrared detectors.

A generalization of metal-dielectric photonic bandgap materials are metamaterial-containing multilayers. Due to the effect of phase compensation [226], higher and wider bandgaps appear in metamaterial-containing multilayer compared to the conventional all-dielectric and metal-dielectric quarterwave stacks. Such layers exhibit omnidirectional reflection [256, 257]. Like in the case of AR structures based on such multilayers, a problem are large absorption losses.

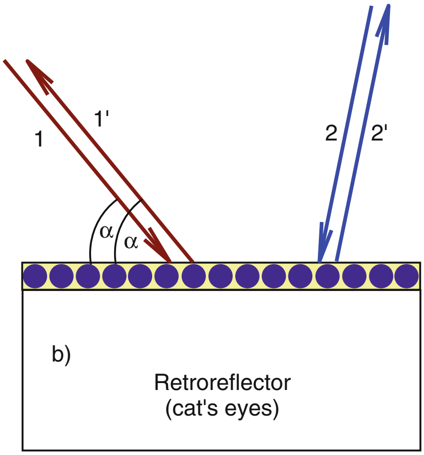


Fig. 2.40 Representation of a retroreflector

2.9.7 Retroreflectors

In certain types of mirror surfaces the reflected beam is returned in the same direction from which the incident beam arrived. This is maintained over a wide range of directions of the incident beam (Fig. 2.40). The phenomenon is called retroreflection [258]. Retroreflectors are various kinds of catadioptric structures ("cat's eyes"). Basically, a retroreflector may be a surface-diffractive structure with triangular, pyramidal, spherical, etc. shapes of the grating. Incident light typically reflects twice before returning in the direction it came from.

The patterns may also be comparable to the operating wavelength or larger, even much larger, and actually many retroreflector structures in everyday life are macroscopic. However, retroreflectors can also have subwavelength dimensions [259].

Instead of the conventional geometrical optics approach, one can use instead the transformation optics [243] which allows one to tailor the beam path at will. In this way one use the mapping of the optical space to arrive at optical components with vastly improved characteristics, in this case a retroreflector with omnidirectional operation [260].

2.9.8 Diffuse Mirrors

A light beam reflected from a diffuse surface does not remain in the same plane, but may be instead reflected omnidirectionally, i.e., in any spatial direction (e.g., [261]). Such behavior is ideally described by Lambert's cosine law.

Diffuse reflectivity is caused by scattering centers which may be either on the reflecting surface or within the volume of the reflecting body and are usually subwavelength or mesoscopic. Multiple scattering tends to produce a diffusely

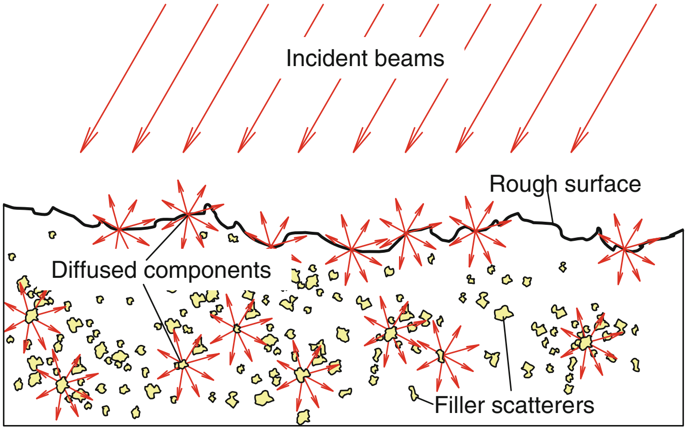


Fig. 2.41 Generalized 2-layer scattering surface

reflected beam emanating in many directions. Two main types of scattering centers are shown in Fig. 2.41.

The total reflectance of a surface is the result of a combination of multiple scattering from the volume inclusions and from the reflector surface. These two components may be in any ratio. Multiple scattering by pigment particles within the coating volume readily achieve diffuse reflectances of 0.80–0.90. The values are much higher if the structure is optimized for diffuse reflectance [262]. Such optimization can be obtained for example by numerical techniques.

One type of diffusing surfaces is a stochastically corrugated reflecting surface. Such structures can range from simple etched metal surfaces to more complex metal-multilayer dielectric stack systems. Surface corrugation is relatively the simplest approach to the fabrication of diffuse reflectors. Typically an optical surface is etched to obtain convenient roughness and then a high-reflection coating is deposited, or, alternatively, a mirror surface is directly roughened.

The other types of diffusing surfaces are various random distribution structures with enhanced scattering in the coating volume. Such structures are macroporous media with pores of the order of hundreds of nanometers (a high refractive index material with air spheres or other kinds of porosity, the “scattering sponges”), colloidal structures (particles of high-refractive index material or dye dispersed in a transparent medium), various powder complexes (sintered or processed otherwise) and generally disordered optical media. A property of all of the volume structures is that their refractive index is strongly—and stochastically or pseudostochastically—dependent on spatial coordinate. In all cases of strongly diffusing surfaces, regardless of their type, it is desirable that the dimensions of the disorder are comparable to the wavelength of light or smaller.

Enhanced backscattering can lead to near-ideal Lambertian illumination and “super-white” coatings, or “superdiffusers”. Diffuse reflectors are described by the theory of disordered media. Generally, this theory is very complex and the methods for its solution include Mie theory, but also various methods like Monte Carlo approach [263].

In structures where light is randomly scattered the scattering strength is characterized by the mean free path, defined as the distance light travels in average between two scatterings. If the mean free path is comparable to the operating wavelength or even smaller than it, coherent backscattering (weak localization) occurs, described as a precursor to Anderson localization (strong localization) [264]. This is similar to the behavior in a photonic crystal, where within its photonic band gap light propagation is also inhibited, but in this case it is caused by the structure disorder, not by its periodicity.

Light propagating through disordered media is multiply scattered, thus its exact description is very difficult. If all interference effects are neglected, which is an extremely rough approximation, then the light transport may be described by diffusion. Light propagates by randomly changing its direction and the characteristics of such transport is the mean free path and thus the diffusion constant. Internal reflections at the boundaries of such medium are also described by diffusion. The details of the calculations of transmission and reflection based on the diffusion approach can be found for example in [265].

Real reflectance can be described analytically or empirically [261]. There are several analytic reflectance models that can be used to describe various types of surfaces. The simplest diffuse source is Lambertian. The Mie theory can compute light scattering by spherical particles, as well as some other simple shapes like elongated ellipsoids [266].

2.10 TIR Structures

One of the approaches to the increase of the optical path through the active area is to use total internal reflection (TIR) and thus ensure multiple reflections from within the detector. To this purpose the active area surface is roughened (textured) in some way, for instance by etching. The obtained roughness may be stochastic, or alternatively some kind of regular pattern may be formed. The pattern shape ensures that the photon flux, after entering the detector and reflecting from the highly reflective backside, arrives from the inside to the corrugation surface at an angle larger than Brewster's. The goal is to obtain total internal reflection of as large percentage of the returning flux as possible. The result is that the absorption path within the active region may be increased several times.

The concept of TIR structures and that of antireflection DOEs are connected. In both of those similar or even identical surface reliefs are used. The difference is in their dimensions. While in the case of AR reliefs the texture constant remains subwavelength ($\Lambda < \lambda$ or in the worst case $\Lambda \approx \lambda$—in the case when $\Lambda \ll \lambda$ the effective medium approximation is valid), in the case of TIR reliefs $\Lambda > \lambda$ or $\Lambda \gg \lambda$—in this text we assume that TIR structures are those where interference effects

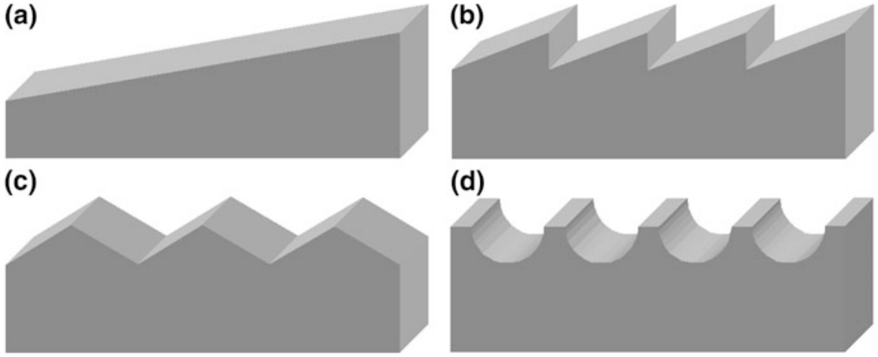


Fig. 2.42 1-D profiles of TIR structures for IR detectors; **a** wedge detector; **b** sawtooth structure; **c** V-groove (triangular structure); **d** hemispherical holes

can be neglected, or alternatively, those with features much larger than mesoscopic and subwavelength ones.

All nonbinary geometries listed in the text on AR subwavelength DOEs can be utilized for TIR reliefs as well: sawtooth geometries (triangles with one vertical side) [267], triangular (and their special case, V-grooves [268]), sinusoidal, hemispherical [269] and irregular corrugations (this includes stochastically distributed tetrahedrons with different dimensions [201, 270], etc. All of these geometries may be 1D or 2D. Figure 2.42 shows examples for some of the mentioned TIR geometries.

In 1970 Lee and Sze described a wedge-shaped TIR structure [271] to improve performance of Si diodes. This was a planar structure polished to obtain a sloped active area (i.e., a sawtooth detector with its period $\Lambda/2$ equal to the full length of detector) and thus receiving incident radiation at an oblique angle. One of the solutions to follow was sawtooth geometry, and another subdividing the structure to segments to reduce it to planar geometry [272] (similar to the connection between a conventional refractive lens and the corresponding Fresnel lens).

V-groove TIR structures (or, as a variation, a truncated—trapezoidal shape) and tetrahedral ones were successfully utilized for solar cells [200, 267].

Several different microfabrication techniques are used to form TIR structures. One of these is photolithography combined with bulk micromachining [196], also ion etching and mechanical polishing [271]. An application of isotropic etching is presented in [269]. TIR structures may be fabricated on one side of the detector only or on both of its sides and may be combined with, e.g., AR films/structures.

The calculation of quantum efficiency improvement by TIR is done by standard methods of geometrical optics (ray tracing) [268]. The wedge-shaped detector was calculated in [271], while triangular and sawtooth structures were analyzed in [268]. A typical approach is to divide the optical surface into a network of points where reflection and refraction of different beams is determined.

2.11 Resonant Cavity Enhancement (RCE)

Absorption within detector can be significantly increased by placing the active area into a resonant cavity, e.g., in a Fabry-Perot resonator. If the distance between resonator mirrors is equal to a small number of half wavelengths and if the product of the effective absorption coefficient and the active area thickness is small enough, the cavity will maintain a single allowed optical mode, while the other modes will be suppressed. In that manner interference causes a large change of optical mode distribution in comparison to the situation in conventional detector structures, including a change of spontaneous emission [273]. Detectors in a 1D optical resonant cavity are denoted as Resonant Cavity Enhanced (RCE) [251] or REsonant CAvity Photodetector (RECAP) [274].

The amplification of the radiation at the resonant wavelength in an RCE device and rejection of others results in a increased spectral selectivity of resonant detectors. A logical consequence is an increase of the detector speed [275] which becomes limited by carrier transit time. Another consequence is an increase of the allowed operating temperature because of the shift of the BLIP limit [276]. The high spectral selectivity of RCE detectors may prove advantageous or disadvantageous, depending on the concrete application.

Interference structures with multiple reflection applied in photodetection were analyzed as early as in 1970s [277]. At the end of 1980s and at the beginning of 1990s the interest for these detectors suddenly increased. It was the result of the requirements for faster photodetectors with narrow spectral range in optical telecommunication, i.e., with minimal crosstalk in wavelength division demultiplexers (WDM). This coincided with the availability of the convenient fabrication technologies [243].

RCE structures were used to improve performance of different optoelectronic devices—Schottky diodes [278, 279], p-i-n devices [243], avalanche photodiodes [280], MSM detectors [281], phototransistors [282, 283], LED diodes [284], etc. In the infrared range RCE was used for p-i-n diodes [285], photoconductors [286] and LED diodes [287].

Figure 2.43 shows a typical RCE photodetector structure.

The absorptive active region (for MWIR and LWIR devices typically fabricated from a narrow-bandgap material, its thickness d and the absorption coefficient α) is positioned between two layers of wider-bandgap material (thickness L_1 and L_2 and the absorption coefficient of both layers α_{ex}). DBR mirrors are on the top and the bottom side. In some cases the top Bragg mirror does not exist, and its role is assumed by the interface between semiconductor and the incident medium (air), which furnishes a top surface reflection coefficient of about 30 % [251].

The consideration of an RCE photodetector reduces to an analysis of a 1D microcavity with losses (absorption used for the detection). The interfaces between the DBR mirrors and the wide-bandgap layers are described by the Fresnel complex reflectances $r_1 = |r_1|\exp(i\psi_1)$ and $r_2 = |r_2|\exp(i\psi_2)$.

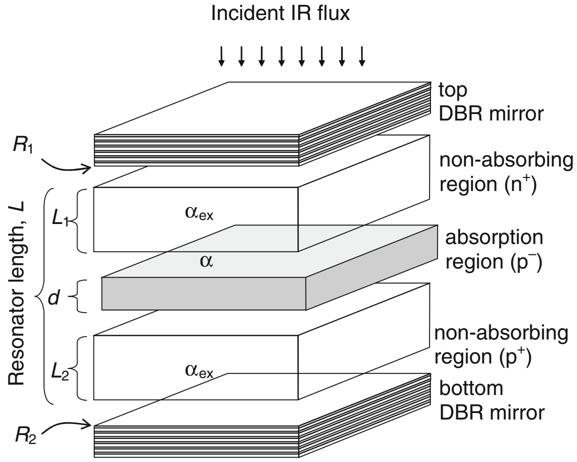


Fig. 2.43 Schematic presentation of an RCE photodetector

If the refractive index of the active region is denoted by n and the wave propagation constant is $\beta = 2n\pi/\lambda$, the quantum efficiency of the photodetector can be obtained as the ratio of the absorbed optical power to the incident power [251], taking into account the interference effects and neglecting the spatial distribution of the optical field within the RCE structure

$$\eta = \left\{ \frac{e^{-\alpha_{\mathrm{ex}}L_1} + e^{-\alpha_{\mathrm{ex}}L_2}|r_2|^2 e^{-\alpha_{\mathrm{sve}}L}}{1 - 2r_1 r_2 e^{-\alpha_{\mathrm{sve}}L}\cos(2\beta L + \psi_1 + \psi_2) + |r_1|^2 |r_2|^2 e^{-2\alpha_{\mathrm{sve}}L}} \right\} \left(1 - |r_1|^2\right)\left(1 - e^{-\alpha d}\right), \tag{2.34}$$

i.e.,

$$\eta = A\left(1 - |r_1|^2\right)\left(1 - e^{-\alpha d}\right). \tag{2.35}$$

Here $\alpha_{\mathrm{sve}} = [\alpha_{\mathrm{ex}}(\mathrm{L}_1 + \mathrm{L}_2) + \alpha d]/L$. The factor A describes the enhancement of quantum efficiency by interference effects. The peaks of quantum efficiency are obtained when the term in the cosine is equal to an integer number of 2π. The wavelength interval between two peaks is denoted as free spectral range.

The spatial distribution within a resonant cavity can be calculated by introducing the standing wave (SW) factor with a position-dependent value. In that case the absorption coefficient within the active area can be replaced by its effective value [251]

$$\alpha_{\mathrm{eff}} = \mathrm{SW}\alpha = \frac{\frac{1}{d}\int_0^d \alpha(Z)|E(z,\lambda)|^2 \mathrm{d}z}{\frac{2}{\lambda}\int_0^{\lambda/2} |E(z)|^2 \mathrm{d}z}. \tag{2.36}$$

If an assumption is made that the product of the absorption and the layer thickness is small, that the absorption coefficient and the refractive index are constant and that the reflection at the boundary of the active layer can be neglected, the following approximation is valid

$$|E|^2 = \frac{\left|1 - r_1^2\right|}{\left|1 - r_1 r_2 e^{i(2\beta L + \psi_1 + \psi_2)}\right|^2}\left[1 + r_2^2 + 2r_2 \cos(2\beta(L - z) + \psi_2)\right]|E_{\mathrm{dir}}|^2 \quad (2.37)$$

where E_{dir} is the electric field corresponding to the incident wave. Then the standing wave coefficient is

$$\mathrm{SW} = 1 + \frac{2r_2}{\beta d\left(1 + r_2^2\right)}[\sin \beta d \cos(2\beta L_2 + \beta d + \psi_2)] \quad (2.38)$$

As an illustration, we consider here resonant cavity enhancement of a HgCdTe detector at 10.6 μm. If we choose Ge–PbF_2 material pair for the RCE Bragg mirrors (Ge thickness 0.66 μm, PbF_2 thickness 1.63 μm), the structure shown in Fig. 2.12 can be designed in the following manner: the top mirror is PbF_2–Ge–PbF_2–Ge–PbF_2, followed by a Ge defect with a thickness of 2.12 μm. The nonabsorbing region is CdTe 0.38 μm, followed by the $Hg_{1-x}Cd_xTe$ active region. The bottom side nonabsorbing region is CdTe 0.76 μm, and the bottom mirror is four Ge–PbF_2 pairs followed by a single Ge layer.

The cadmium molar fraction of the photodetector was $x = 0.165$, which is the optimum for the operation at 300 K. The detector-active area was assumed to be 1 mm x 1 mm, one of the standard dimensions for uncooled HgCdTe photoconductors for the detection of CO_2 laser radiation. The bias voltage was 1.5 V, and the detector resistance 50 Ω.

Figure 2.44 shows the calculated reflection coefficient from the incident medium to the active area. R_1 denotes the top (front) surface, and R_2 bottom, as seen from the active area.

Figure 2.45 shows the dependence of quantum efficiency on the active region thickness. The same diagram also shows this dependence for a conventional HgCdTe photodetector ($R_1 = 0$). It was assumed that the detector had a backside Bragg mirror with its parameters identical to the one used in the RCE case. The front surface was protected with a PbF_2 antireflection coating. The diagram also shows the dependence of the similar structure with taking into account interference effects, i.e., an unoptimized RCE structure.

The first maximum in the RCE dependence in Fig. 2.45 is found for a thickness of 0.184 μm, and its value is 98 %. The spectral dependence of this quantum efficiency is shown in Fig. 2.46.

Figure 2.47 shows the changes of the quantum efficiency with varying thickness of the active region. This “sensitivity” of detector parameters to the variations of the thickness of the active region is important from the technological point of view, since a more “spread” curve means more relaxed conditions for the fabrication of

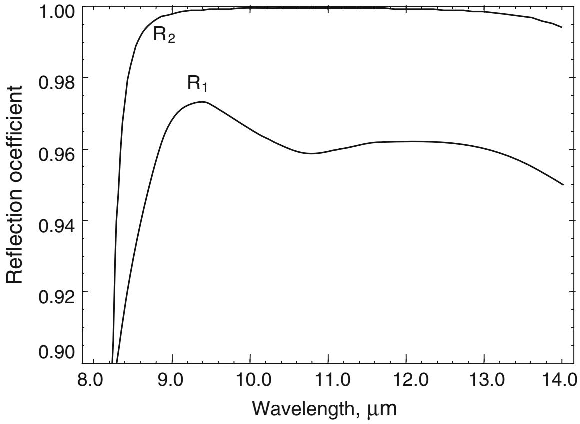


Fig. 2.44 Reflection coefficients of top and bottom DBR layers of an RCE structure (as seen from the active region)

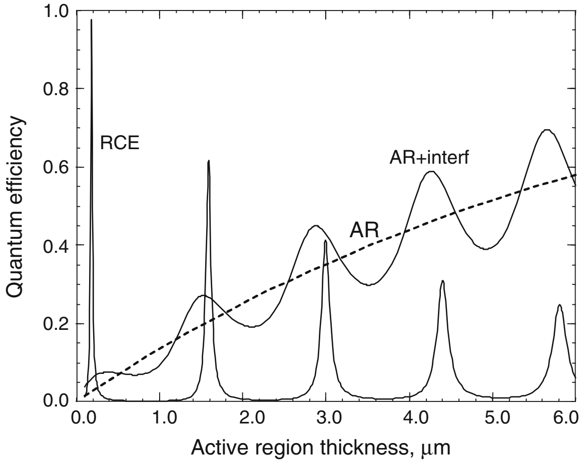


Fig. 2.45 Quantum efficiency of HgCdTe RCE photodetector designed for the LWIR wavelength of 10.6 µm. *RCE* resonant structure; *AR* (*dashed*) conventional detector; *AR* + *interf* detector with an antireflection layer taking into account interference (unoptimized RCE)

the RCE structure. It can be seen that the thickness changes of about 3 nm result in a quantum efficiency drop of near 10 %, while an error of 6 nm results in a 20 % drop.

Figure 2.48 shows spectral responsivity and detectivity of the HgCdTe RCE detector with an active area thickness of 0.184 µm.

Fig. 2.46 Quantum efficiency for an HgCdTe RCE detector 0.184 μm thick

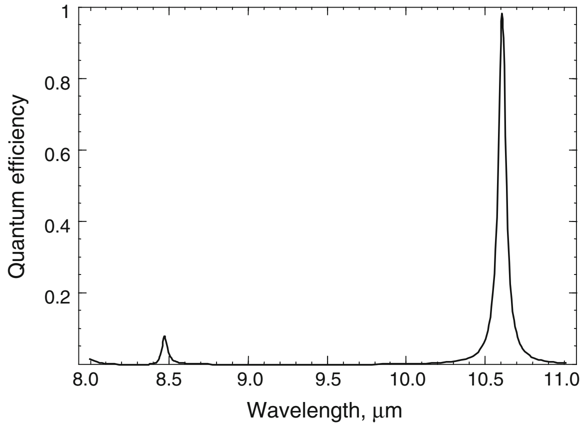
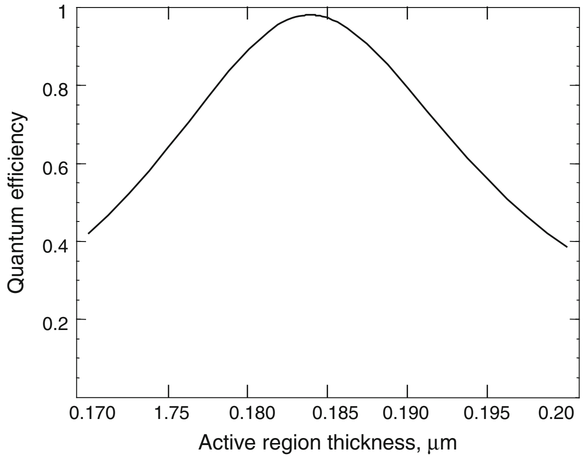


Fig. 2.47 Quantum efficiency of a HgCdTe RCE photodetector versus active region thickness



The maximum current responsivity (at 10.6 μm) of the 0.184 μm thick structure is 0.129 A/W, and the corresponding specific detectivity 1.94×10^9 cmHz$^{1/2}$/W. For a comparison, a conventional structure reaches the same responsivity for an active region thickness of 27.4 μm, and an unoptimized RCE for 14 μm. Also for a comparison, the best conventional uncooled detectors with d = 5 μm [8] reach $\eta \approx 0.60$.

A conventional uncooled detector (no interference, but the same Cd molar fraction, the same back mirror and an AR coating) with a thickness identical to the RCE (i.e., 0.184 μm) reaches $R_I = 0.034$ A/W and thus $D^* = 5.2 \times 10^7$ cmHz$^{1/2}$/W, i.e., an almost two orders of magnitude lower value. A conventional detector with a current responsivity identical to that of the RCE (d = 27.4 μm) reaches $D^* = 1.59 \times 10^8$ cmHz$^{1/2}$/W. This means that the specific detectivity is more than 12 times higher if all other parameters are identical.

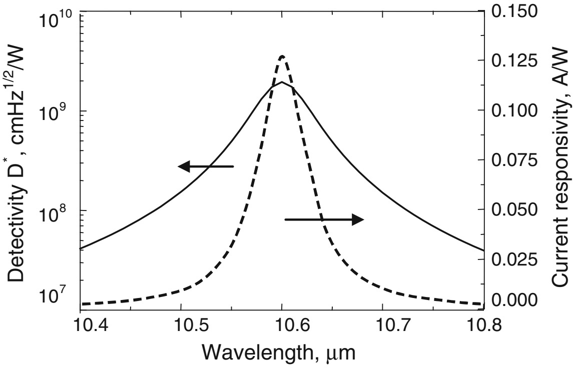


Fig. 2.48 Spectral current responsivity and specific detectivity of a HgCdTe RCE detector optimized for 10.6 μm

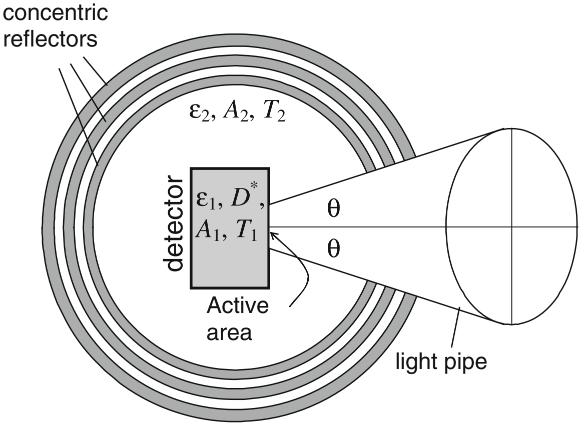


Fig. 2.49 Performance improvement by radiative shield

2.12 Radiative Shields

Radiative shields are screens with low emissivity (high-reflection coefficient) partially or completely enclosing the detector. In 1981 Putley proposed this method to improve BLIP detectivity of an ideal uncooled thermal detector [6]. We consider here the use of a radiative shield as a macro- or microcavity that may be resonant or nonresonant. In the most general case it would be a full 3D photonic bandgap structure (the PBG coinciding or being wider than the electronic bandgap of the semiconductor material used). A defect mode/light pipe may be introduced into the radiative shield on the incident side.

Figure 2.49 shows a schematic presentation of a photodetector surrounded by a radiative shield. A light pipe is shown on the right side as a cone-shaped opening with a solid angle θ. Similar to the radiative shield itself, its walls are made of highly reflective material. The detector radiates as a blackbody source (temperature T_1) and at the same time its radiation must be in equilibrium with its surroundings.

If the emissivity of the radiative shield ε_2 is much lower than unity, the level of equilibrium radiation emitted by the detector must drop (suppression of spontaneous radiation) and thus its BLIP detectivity will increase. The ideal result would be obtained for an aperture of the light pipe close to zero, but in that case diffractive effect completely change the behavior of the whole system.

Putley proposed to use the so-called superisolators as the material of the radiative shield. These materials were developed for thermal isolation in cryogenic technique and their thermal impedance is larger than that of vacuum. These are actually concentric layers of isolators and highly reflective metal layers.

In our consideration we assume that the detector and the radiative shield are at the same temperature $T_1 = T_2 = T$), which is near or equal to the room temperature. The detectivity of a photoconductive BLIP detector is given by [5]

$$D^*_{\mathrm{PC}} = \frac{\lambda}{2\mathrm{hc}}\sqrt{\frac{1}{\Phi_{\mathrm{b}}}}. \tag{2.39}$$

while for a photovoltaic detector it is $\sqrt{2}$ times larger. Here Φ_{b} denotes the background flux which can be absorbed in a detector with a cutoff wavelength of λ_{co} and which is given by the Planck integral

$$\Phi_{\mathrm{b}} = \frac{1}{\varepsilon_1}\int_0^{\lambda_{\mathrm{co}}} \frac{2\pi c}{\lambda^4}\frac{d\lambda}{e^{\mathrm{hc}/\lambda k_b T} - 1}. \tag{2.40}$$

The factor $1/\varepsilon_1$ takes into account the detector emissivity.

The following expression is valid for the radiation flux of a detector fully surrounded by a radiative shield

$$\Phi_{\mathrm{shield}} = \Phi_{\mathrm{b}}\sqrt{\frac{\varepsilon'}{\varepsilon_1}}, \tag{2.41}$$

While deriving the above equation the angular dependence of emissivity was not taken into account. ε' denotes the equivalent emissivity of the shield. Its value is obtain by a thermodynamical consideration of heat exchange between the detector and the shield. For a diffusively reflective external surface

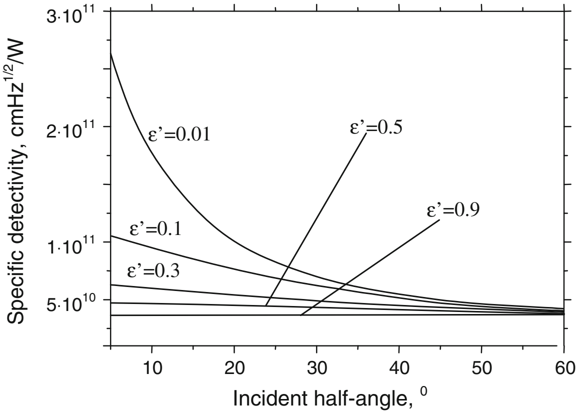
$$\varepsilon' = \left[\varepsilon_1^{-1} + (A_1/A_2)\left(\varepsilon_2^{-1} - 1\right)\right]^{-1} \tag{2.42}$$

and for a specular one

$$\varepsilon' = \left[\varepsilon_1^{-1} + \left(\varepsilon_2^{-1} - 1\right)\right]^{-1} \tag{2.43}$$

In both cases for a low ε_1 and a high ε_2 it will be $\varepsilon' \approx \varepsilon_2$ (we assume that $A_1 \approx A_2$, i.e., that the shield is directly deposited to the active area).

Fig. 2.50 Specific detectivity of photoconductor versus incident half-angle of light pipe for different shield emissivities. $T = 297$ K, $\varepsilon_1 = 0.9$, $\lambda_{co} = 10.6$ μm



If there is a light pipe, the above expression for the Φ_{shield} is modified my multiplying it with $\sin^2(\pi - \theta)$, while the flux arriving through the light pipe is given by $\Phi_{\text{pipe}} = \Phi_b \sin^2\theta$. The total flux is obtained by summing these two fluxes, which is equivalent to summing the noise equivalent powers for each flux separately

$$\Phi_{\text{tot}} = \left[\frac{\varepsilon'}{\varepsilon_1^2}\cos^2\theta - \frac{\sin^2\theta}{\varepsilon_1}\right]\int_0^{\lambda_{\text{co}}}\frac{2\pi c}{\lambda_1^4}\frac{d\lambda_1}{e^{\text{hc}/\lambda_1 k_b T}-1}. \tag{2.44}$$

We assume that ε_1 and ε_2 do not depend on wavelength in the spectral range of interest.

Figure 2.50 shows the influence of the decrease of the emissivity of the radiative shield and of the decrease of the solid angle of the light pipe to the increase of the maximum specific detectivity of the detector. Figure 2.51 gives a comparison of a detector without a radiative shield, the one with the shield having a light pipe with a 15° angle (i.e., relatively large) and the one with a near-ideal case of radiative shield (light pipe angle tends to zero). Finally, Fig. 2.52 shows the specific detectivity of a photoconductor versus wavelength for a total enclosure within a radiative shield for various shield emissivities.

2.13 Photonic Crystal Enhancement (PCE)

Photonic crystals or photonic bandgap materials may be defined as artificial 1D, 2D, or 3D structures where the dielectric constant periodically changes, with a result that photons within it behave similarly to electrons in semiconductor crystal lattice [288, 289]. A photonic bandgap (PBG) appears in photonic crystals, an energy

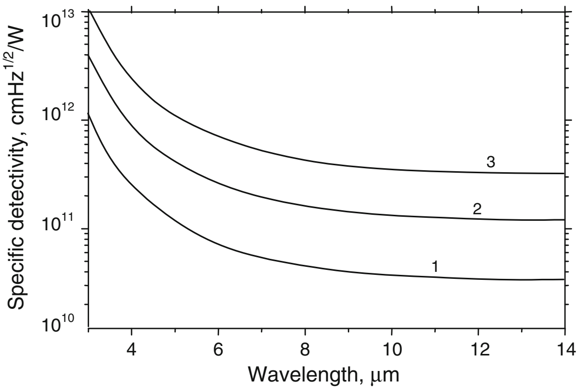


Fig. 2.51 Spectral dependence of specific detectivity of photoconductor for *1* conventional detector (no radiative shield); *2* radiative shield with a light pipe incident angle of θ = 15°; *3* the case of ideal total radiative shield, θ = 0°. $T = 297$ K, $\varepsilon_1 = 0.9$, $\varepsilon_2 = 0.01$, $\lambda_{co} = 10.6$ μm

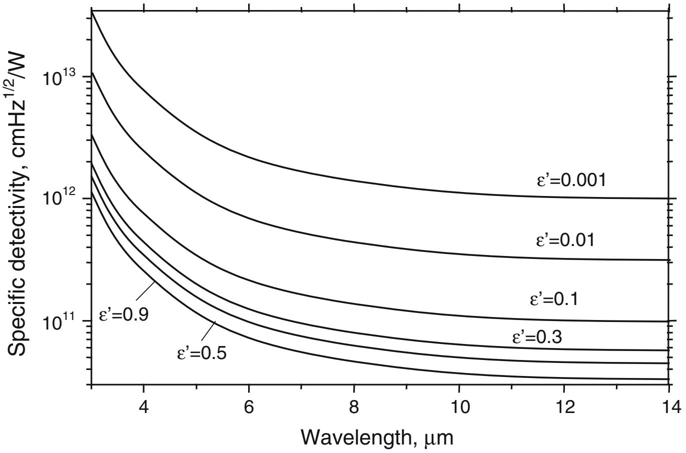


Fig. 2.52 Specific detectivity of a photoconductor versus wavelength for a total enclosure within a radiative shield for various shield emissivities. $T = 297$ K, $\varepsilon_1 = 0.9$

range in which electromagnetic waves cannot propagate through the structure. A photonic crystal behaves as a high reflection coefficient mirror for all wavelengths within the photonic bandgap, and the waves with such wavelengths are evanescent within it. In 3D structures this behavior is independent on the propagation direction (i.e., the radiation incident angle), while in 2D structures it has a strong angular dependence in a plane orthogonal to the propagation direction, and in 1D structures in two planes, which are at the same time orthogonal to the propagation direction and one to each other.

The fabrication of 3D photonic crystals for optical wavelengths is still a formidable technological problem, since their structure must have mesoscopic features (comparable to the operating wavelength radiation), and their fabrication accuracy must be at least an order of magnitude better. A number of methods have been proposed for their fabrication [289]. The problem is much simpler with 2D structures where it requires submicrometer photolithography, while in the case of 1D crystals it reduces to conventional thin film technologies.

An important property of photonic crystals are defect modes, i.e., the allowed energy modes within their photonic bandgap which appear when the symmetry of the photonic crystal is broken [290, 291]. They are analogous to the impurity levels in semiconductor lattices. The symmetry can be perturbed by increasing the dimensions of the high-refractive index material in the photonic unit cell (donor, introducing a mode originating from the bottom of the conduction band) or by removing a part of it (equivalent to adding the low-index material; this is acceptor, originating from the top of the photonic valence band). The result in both cases is a sharp peak at the defect wavelength. The defect mode wavelength can be adjusted by changing the properties of the defect (its thickness or the refractive index).

The use of photonic crystals for the enhancement of solar cells by utilizing PBG structure instead of the high-reflection Bragg mirrors has been reported by Bermel et al. [292]. They proved that a six-period triangular two-dimensional PBG structure made of air holes in silicon increases power generation more than 2 %, even more if a combination of a Bragg mirror and 2D PBG is used. A similar improvement is obtained if an eight-period inverted opal photonic crystal [293] is utilized.

The use of photonic crystal cavity to improve characteristics of infrared photodetectors (PCE, Photonic Crystal Enhancement) was described in [237]. The photodetector is enclosed within a PBG structure (3D in a general case) with a photonic bandgap either equivalent to the electronic bandgap of the utilized semiconductor material or larger than it. A defect may be introduced in the photonic crystal on the incident side of the photodetector, and its transmission peak adjusted to a desired wavelength. Such an approach improves the performance of the detector in two ways at the same time. As other methods of photodetector enhancement through optical path increase, it increases the probability of absorption of useful radiation in the active region and thus improves quantum efficiency and the responsivity. At the same time, reabsorption (photon recycling) increases the radiative lifetime and thus shifts the BLIP detectivity limit. The basic concept of the method is illustrated in Fig. 2.53. Although the presented photonic crystal is an inverted opal, any other type of PBG structure may be actually used instead.

Both RCE and PCE represent cavity enhancement methods. The main difference between them is that in the latter the cavity maintains more than a single mode, thus ensuring a range of operating wavelengths within the photonic bandgap.

The influence of a PBG structure to the detectivity of a photonic detector may be considered in a manner analogous to that presented in Sect. 2.12. Actually a photonic crystal may be considered the ideal case of a radiative shields, describing

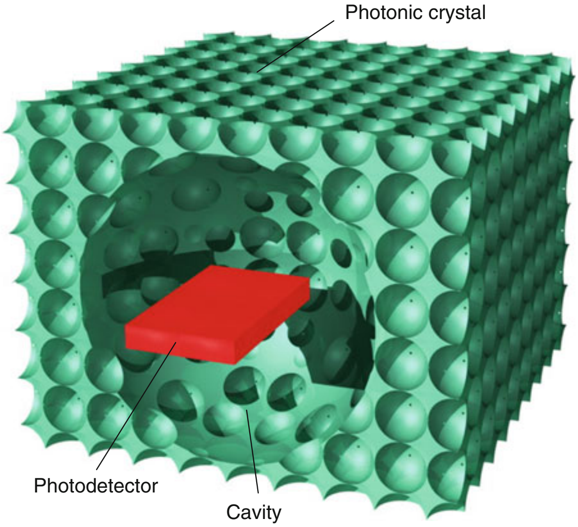


Fig. 2.53 Schematic presentation of cavity enhancement in a photodetector using photonic crystal. The *front half* of the PBG structure is removed to show the photodetector slab. The photonic crystal is an inverted opal, but this is for illustration purpose only since it can be any other type of PBG structure

the situation when the spatial angle of the light pipe is zero. The BLIP detectivity of an uncooled photoconductor calculated utilizing the approach from 9.3 is shown in Fig. 2.53.

In the further text, we shortly consider PCE utilizing a 1D photonic crystal with a defect. This situation is chosen for its simplicity. It can be generalized to 2D and 3D structures, but qualitatively the results remain similar.

In 1D case a defectless photonic crystal reduces to a Bragg distributed mirror (DBR). The defect modifies its spectral characteristics by introducing a transmission maximum dependent on the defect properties and deforms the 1D photonic bands to some extent.

Figure 2.54 shows a photodetector enclosed within a 1D PBG structure.

Using the Maxwell equations for the general case of propagation of electromagnetic waves through a one-dimensional medium, a wave equation can be written [179] with a form equivalent to the Schrödinger equation for electrons in semiconductor crystal in 1D case

$$\frac{\partial^2 E(x,t)}{\partial x^2} = \frac{\varepsilon(x)\partial^2 E(x,t)}{c^2 \quad \partial t^2}, \tag{2.45}$$

and if we assume that the electric field has the form of a plane wave $E(x,t) = E(x)\, e^{-i\omega t}$:

$$\frac{\partial^2 E(x)}{\partial x^2} = -\frac{\omega^2}{c^2}\varepsilon(x)E(x). \tag{2.46}$$

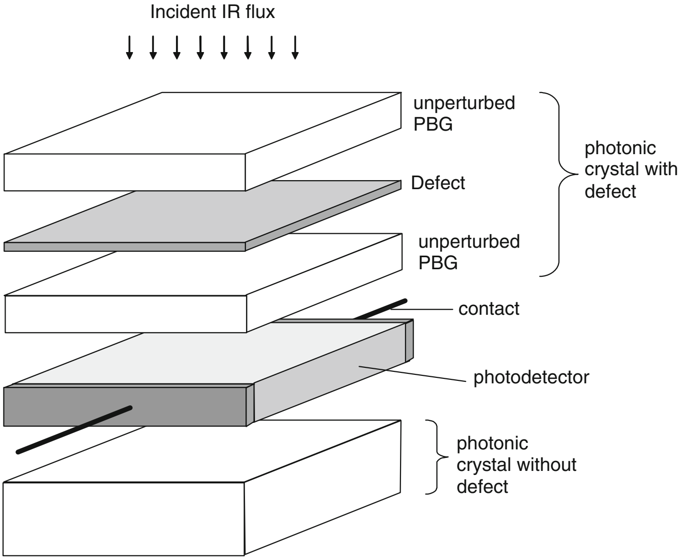


Fig. 2.54 Idealized presentation of a photodetector surrounded by 1D photonic crystals. The photonic bandgap of both of the crystals coincides with the spectral sensitivity range of the detector. Additionally, the photonic crystal at the incident side has a built-in defect at the desired operational wavelength

Here the dielectric permittivity is spatially periodic, $\varepsilon(x + d) = \varepsilon(x)$, which means that the Kronig-Penney model is applicable. If we introduce the Bloch vector, i.e., the crystal momentum κ, the periodicity of the electric field will be described by $E(x+d) = e^{i\mu d}E(x)$—Bloch or Floquet condition. The solution of the wave equation for an infinite 1D lattice with periodically changing dielectric permittivity should have the form of a sum of a direct and a reflected wave

$$E_i(x) = A_i e^{ik_i x} + B_i e^{-ik_i x}, \tag{2.47}$$

where $k_i = \omega\varepsilon_i^{1/2}$. While solving the above equation, we assume the continuity of the electric and the magnetic field on the boundary between layers. For the photonic bandgap of a 1D photonic crystal, we obtain the following transcendent equation

$$\cos(\kappa d) = \cos[k_1(d-a)]\cos(k_2 a) - \frac{1}{2}\left(\frac{k_1}{k_2} - \frac{k_2}{k_1}\right)\sin[k_1(d-a)]\sin(k_2 a), \tag{2.48}$$

This approach for the 1D case reduces to the consideration of a Bragg dielectric mirror.

In the case when there is a defect, the wavelength of its localized peak is obtained by solving the equation

$$tg[k_1(d-a)] = -\frac{tg(k_2a)}{2}\left\{\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right) \pm \left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)\frac{\cos(k_1d)}{\cos[k_1(d-a)]}\right\}. \tag{2.49}$$

The equation

$$\sigma\cos(\gamma d) = \cos[k_1(d-a)]\cos(k_2a) - \frac{1}{2}\left(\frac{k_1}{k_2}+\frac{k_2}{k_1}\right)\sin[k_1(d-a)]\sin(k_2a) \tag{2.50}$$

determines the symmetry (the sign factor $\sigma = \pm 1$) and the position ($1/\gamma$) of the defect modes.

Using (2.38) a photonic bandgap map can be obtained for 1D crystals [244]. Figure 2.55 shows two maps for 1D PBG structures. One of them is designed for the atmospheric window (3–5) μm an consists of a pair silicon–silica, while the other is intended for the (8–12) μm window and its material pair is lead fluoride and germanium.

The reflection coefficient for an unperturbed PBG structure chosen according to the map of photonic bandgaps is calculated by the transfer matrix method.

A defect is taken into account by writing the transfer matrix for the part of the PBG structure starting from the incident surface and ending with the layer next to the defect. According to the notation from Sect. 2.9.4, the degree of this transfer matrix is M/2, because usually the middle layer is replaced with a defect. Thus

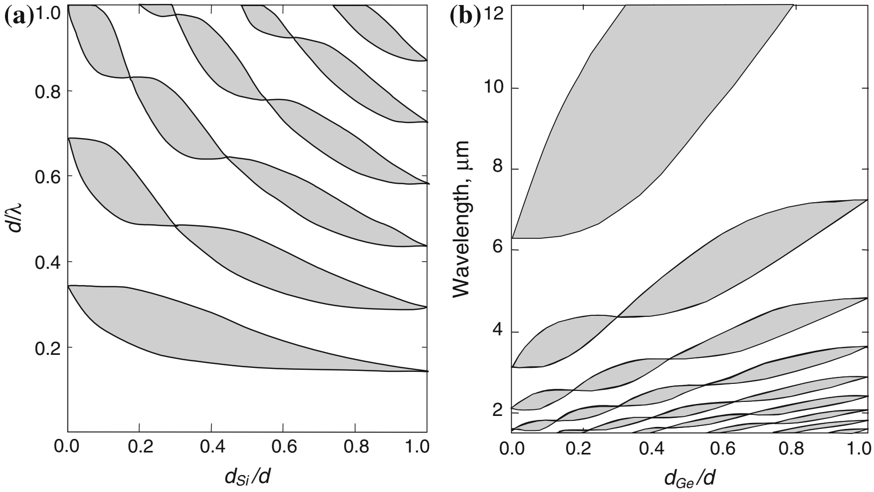


Fig. 2.55 Photonic bandgap maps for 1D photonic crystals for different material pairs. **a** silicon/silica; **b** germanium/lead fluoride. *Shaded areas* correspond to photonic bandgaps

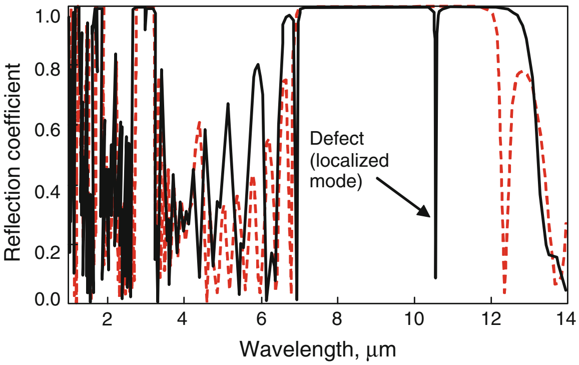


Fig. 2.56 Spectral dependence of the reflection coefficient for 1D Ge/PbF$_2$ photonic crystal with seven periods. *Dashed line* denotes a defectless structure, *solid line* is valid for a structure with a defect at 10.6 µm

obtained matrix is multiplied from the right with the matrix describing the defect layer and then again with a M/2 degree matrix which is identical to the first one. The reflection coefficient is further calculated in the previously described manner.

As an example we consider the design of a 1D photonic crystal for the improvement of the photodetector performance in the LWIR range. According to the map shown in Fig. 2.55b a convenient ratio of the thickness values of the constituents is $d_{Ge}/d = 0.31$, where $d = d_{PbF2} + d_{Ge} = 1.81$ µm, i.e., $d_{Ge} = 0.57$ µm and $d_{PbF2} = 1.24$ µm. For thus chosen PBG structure, the reflection and transmission coefficient are calculated. Then the thickness of the middle Ge layer is varied to determine the transmission peak wavelength (reflection minimum) versus the defect thickness. Using thus obtained value the thickness of the defect is determined which furnishes maximum transmission at the desired wavelength, in this case at 10.6 µm.

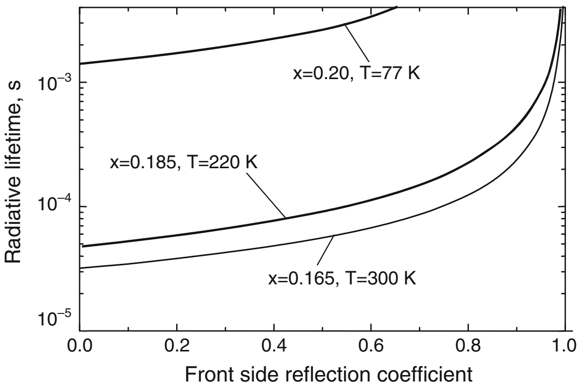
Figure 2.56 shows the calculated transmission of a photonic crystal with 7 Ge/PbF$_2$ layers. The defect at 10.6 µm is obtained by increasing the thickness of the middle layer from 0.57 to 1.545 µm. The transmission peak of the defect mode is narrower if the number of the layers around the defect is larger.

2.14 Radiative Lifetime Increase

It may be generally said that each technique for detector performance enhancement by way of the increase of optical path through the active region simultaneously increases radiative lifetime and thus increases the BLIP detectivity limit. The benefit for detector performance, however, can be seen only in those cases when radiative lifetime is the dominant one. This case is analyzed in this section.

Even the presence of a conventional high-reflection layer (metal, DBR or hybrid mirror) at the photodetector backside influences the reabsorption radiative lifetime

Fig. 2.57 Radiative lifetime versus front-side reflection coefficient for a $Hg_{1-x}Cd_xTe$ detector, thickness 5 μm, at 77 K ($x = 0.20$), 220 K ($x = 0.185$), and 300 K ($x = 0.165$)



[28, 29]. Any strict division between the methods for radiative lifetime increase and those for the optical path increase is impossible.

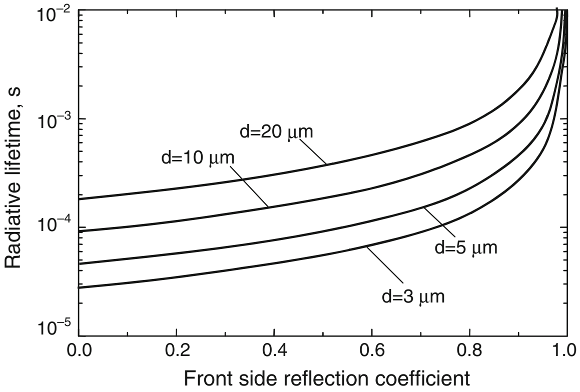
As an illustration, we consider the calculation of radiative lifetime of bulk $Hg_{1-x}Cd_xTe$ detector structures surrounded by Ge/PbF_2 photonic crystal. HgCdTe bandgap was determined according to [25], Fermi level according to [294], and HgCdTe refractive index according to [295].

Figure 2.57 shows radiative lifetime versus front-side reflection coefficient for an idealized $Hg_{1-x}Cd_xTe$ structure mounted on a highly reflective back surface $R_2 \approx 1$. Operating temperatures are 77 K (standard for BLIP photoconductors at 10.6 μm), 220 K (temperature achievable by multistage thermoelectric coolers) and, for comparison, 300 K. Cd molar fraction values are 0.205, 0.185, and 0.165, respectively, which corresponds to sensitivity maxima at 10.6 μm at each of these temperatures.

Figure 2.58 shows the dependence of radiative lifetime on front-side reflection coefficient for $x = 0.185$ and at $T = 220$ K for detector thickness 3, 5, 10, and 20 μm.

Radiative lifetime increases with a larger number of periods of photonic crystals and decreases with operating temperature. From the point of view of radiative

Fig. 2.58 Radiative lifetime versus front-side reflection coefficient for a $Hg_{1-x}Cd_xTe$ detector, thickness 3, 5, 10, and 20 μm at 220 K, composition $x = 0.185$



lifetime the immersion of the photodetector within a PBG structure is equivalent to a decrease of the operating temperature. The performance improvement is larger for a higher refractive index contrast within the photonic crystal and can be obtained in a larger spectral range.

2.15 Plasmonics for Light Trapping

A field of research that is currently exponentially growing is plasmonics [243]. It deals with the propagation of surface plasmons polaritons (SPP) through metal-dielectric structures (or generally, through composite materials consisting of materials with positive and with negative relative dielectric permittivity in the wavelength range of interest).

Surface plasmons polaritons are oscillations of free electrons in conductive material (the negative permittivity part of the composite) coherently coupled with electromagnetic radiation at an interface with dielectric. The ε-negative material is denoted as plasmonic material. Most often the ε-negative material is metal (typical examples are gold and silver). Other plasmonic metals include chromium, copper, alkali metals, various alloys, etc.). Alternative plasmonic materials are transparent conductive oxides (usable in infrared range), semiconductor materials (gallium arsenide is used in mid infrared), graphene, etc. [243].

A SPP is confined to the interface between the positive and the negative permittivity part and is evanescent away from the interface. SPPs can be propagating along the interface, or they can be nonpropagating, i.e., spatially confined to, e.g., a plasmonic nanoparticle (localized surface plasmons polaritons) [243].

Plasmonic nanocomposites can be ordered on 1D, 2D, or 3D level. They can be periodic, quasiperiodic [296], aperiodic [224], or random [204]. The building blocks of these functions themselves may have different shapes, from simple to complex and from regular to irregular [297], Fig. 2.59.

An important disadvantage of SPPs is their resonant nature, which causes a narrow bandwidth of operation. Another one is related to large absorption losses.

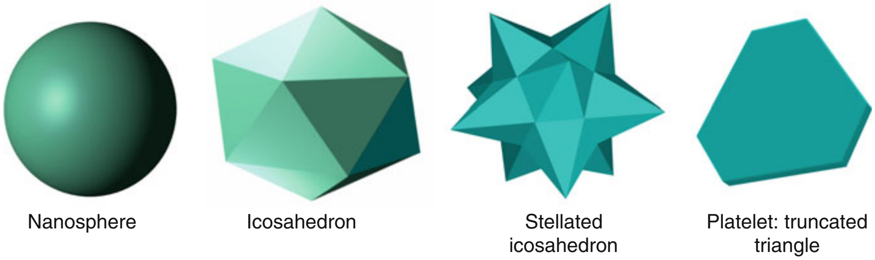


Fig. 2.59 Building blocks of plasmonic nanocomposites: nanoparticles fabricated by bottom-up approach

The relative dielectric permittivity of plasmonic materials is negative below plasma frequency, and its dispersion is by Drude electron resonance model [298]

$$\varepsilon = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)}, \tag{2.51}$$

where ω_p is the plasma frequency, Γ is the damping factor describing losses, and ε_∞ is the asymptotic relative dielectric permittivity. The plasma frequency is determined by the properties of free carriers as

$$\omega_p^2 = \frac{n_e e^2}{m^* \varepsilon_0}, \tag{2.52}$$

where n_e is electron concentration, e is the free electron charge (1.6×10^{-19} C), ε_0 is the free space (vacuum) permittivity (8.854×10^{-12} F/m), and m^* is the electron effective mass. The damping factor is

$$\Gamma = \frac{e}{\mu m^*} \tag{2.53}$$

where μ is mobility of free carriers.

Dispersion relation of plasmonic material can be tuned by design, leading to an optical behavior surpassing natural materials. These are plasmonic metamaterials [205]. One can use them to obtain superluminal group velocities (“fast light”), near-zero (“slow light”), or negative (“left-handed light”) [207]. This is of crucial importance for photodetector enhancement, since the design of arbitrary frequency dispersion gives a possibility to convert propagating far field modes into spatially localized near-field modes. This ensures strongly increased density of states and thus light trapping in subwavelength volumes.

Elimination of absorption losses that are obviously undesired in photodetection is a field of active investigation, and different schemes were proposed for its avoidance [299]. One of the approaches is the use of alternative plasmonic materials [300] including graphene [301].

The applicability of plasmonics for photodetector enhancement including infrared devices has been recognized in 1970s. The first proposed applications of SPP were in photodetection [243]. Surface plasmon polariton-mediated light trapping schemes for photodetectors [299, 302] may be divided into several distinct groups. Probably the most important is enhanced Mie scattering on plasmonic nanoparticles or nanovoids [303], coupling into guided modes [302], field localization and generation of hotspots near the surface of plasmonic material [304], use of plasmon-based singular optics (optical vortices) [299], use of metamaterial-based transformation optics to map the optical space into a desired shape and with an increased density of states [305] and plasmon-enhanced up-conversion media [306].

Besides nanoparticles and nanovoids as scatterers and nanoantennas, different diffractive structures can be used (gratings, lattices) [129] for field coupling into guided modes; they may be ordered or disordered, as well as subwavelength plasmonic crystals (SPC) which may be periodic [307] or quasiperiodic [243]. Plasmonic nanocomposites with nonresonant operation were also reported [308].

Among subwavelength plasmonic crystals described within the context of infrared photodetection, an important position belongs to 2D arrays of nanoapertures in opaque metal films. Such structures first drew attention for their ability to transmit light in spite of the dimensions of nanoapertures that are much smaller than the operating wavelength. This effect was denoted as extraordinary optical transmission (EOT) [309].

Extraordinary optical transmission is a result of resonant excitation of surface plasmons polaritons at the metal surface. It forces the electromagnetic waves incident to the surface to pass through the apertures. The transmission occurs for all incident electromagnetic energy, not only the one impinging to the EOT aperture. However, the transmitted energy passes only through the apertures. Thus large field localizations appear in the apertures. This behavior effectively corresponds to impedance matching between propagating waves and the perforated metal film. Thus the extraordinary optical transmission arrays act as antireflective structures.

Another application of 2D arrays of subwavelength holes in metal has found an important use in light trapping for infrared detectors. It is based on the so-called designer plasmons.

Pendry et al. [310] have shown that a surface wave propagating along a 2D array of apertures in metal film one can be described by an effective permittivity as

$$\varepsilon_{\text{in-plane}} = \frac{\varepsilon_{\text{hole}}\pi^2 d^2}{8a^2}\left(1 - \frac{\pi^2 c^2}{a^2\omega^2\varepsilon_{\text{hole}}\mu_{\text{hole}}}\right) \tag{2.54}$$

where $\varepsilon_{\text{hole}}$ is the permittivity of the material within the aperture, a is the characteristic dimension of the aperture (side length in the case of square holes), Fig. 2.60, while k_0 is the wavevector in vacuum. The effective plasma frequency of such material is

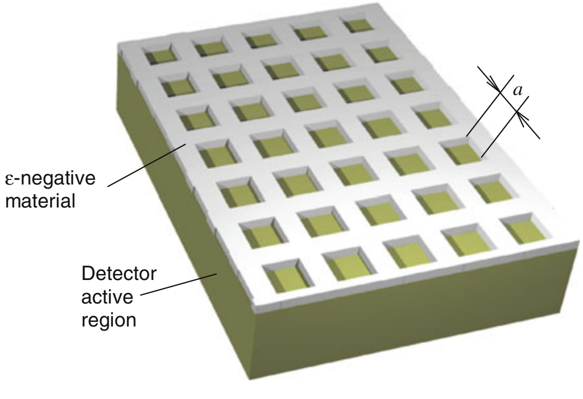


Fig. 2.60 Metallodielectric EOT structure introducing “designer” plasmons with structurally tunable plasma frequency

$$\omega_p = \frac{\pi c}{a\sqrt{\varepsilon_{\text{hole}}\mu_{\text{hole}}}}. \tag{2.55}$$

In other words, the effective dielectric permittivity of an EOT array has a form identical to that of plasmonic materials. At the same time, its plasma frequency and thus its operating range can be freely tuned to the wavelength range of interest. Such surface waves actually mimic real SPP, so that one can fabricate an effective plasmonic material with a desired spectrum. Pendry denoted these surface waves as "designer" plasmons and they are also known as "spoof" plasmons.

The existence of designer SPP is extremely important for infrared photodetectors. One is able to fabricate any desired plasmonic structure and to tune it for the targeted wavelength range. As an example, infrared detectors enhanced by designer plasmon structures tuned to the range of 8–10 μm have been reported [311].

2.16 Optical Antennas

Nanoantenna or optical antenna [243] is a structure that couples with propagating light waves and redirects them into evanescent field. Thus it ensures a highly efficient linking between the two. Depending on the nanoantenna type and its design, the mode confinement in the evanescent field can be tailored and it can be deeply subwavelength. This is of large interest for photodetectors since optical space is basically squeezed into a small region where the interaction with photodetector-active region can be vastly enhanced.

Contrary to the conventional antennas for radio waves, nanoantennas are not connected to any feeding circuitry. They are standalone structures and are used in large arrays. They enhance scattering of electromagnetic waves, the more so the nearer they are to resonance. The literature describes various types of nanoantennas that were experimentally produced. Figure 2.61 shows some of the basic geometries.

The simplest and the most basic nanoantenna is the nanosphere. This structure actually behaves like a dipole. Its scattering properties can be calculated using the Mie theory [266]. Noble metal nanoparticles are often fabricated in spherical form. This makes them the simplest nanoantennas.

If two spherical nanoantennas are brought near to each other so that there is a gap with subwavelength width between them, they form a nanodimer. These nanospheres are electromagnetically coupled, and the coupling gap between them is denoted as the feed gap. A field hotspot appears in the feed gap, where localization is deeply subwavelength and field enhancement is very strong. Much larger field localizations are obtained using such coupled structures than by independent nanospheres. Nanoparticles can be also ordered in an array (nanoparticle chain) [312]. Thus obtained optical antenna effectively behaves as linear nanorod antenna.

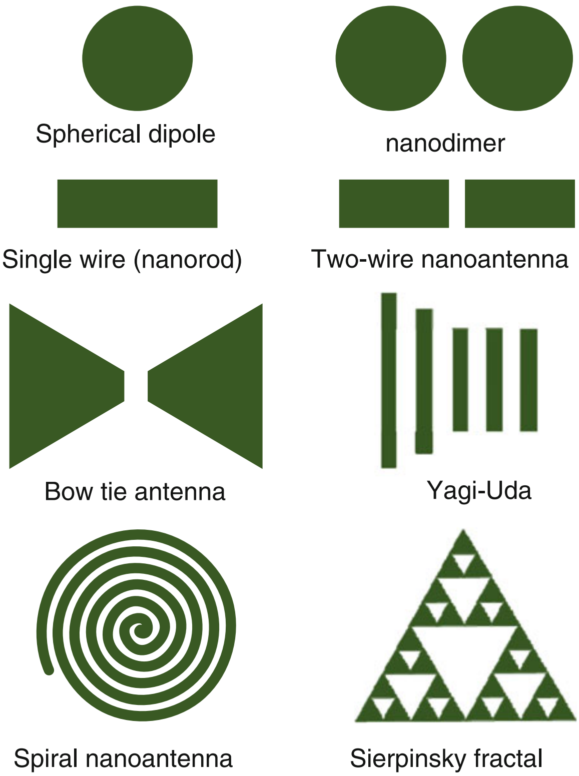


Fig. 2.61 Some different types of experimental plasmonic nanoantennas

A generalization of this case is a nanorod antenna where the dipole length can be tailored. These antennas represent linear dipoles. If two such linear dipoles are aligned and brought together to a subwavelength distance, end-to-end coupling occurs. This is the a two-wire nanoantenna [308]. This can be further generalized by adding two more dipoles perpendicularly to the first ones, all four having a joint feed gap. This configuration is called the cross-antenna.

Another very often met nanoantenna structure is the bowtie nanoantenna [313]. It consists of two triangles aligned along their symmetry axes. A feed gap is formed between their tips. Bowtie antennas have a broader bandwidth together and at the same time ensure large field localizations in the feed gap. A similar type of antenna is the diabolo-type nanoantenna [314], where the triangles overlap.

There is a plasmonic analogy of the well-known Yagi-Uda antenna. It can be fabricated by placing a resonant nanorod antenna between a reflector nanorod and a group of director nanorods [315]. Similar to such antennas used in radiofrequent domain, it ensures a good directivity.

Among the nanoantennas mentioned in the literature are spiral nanoantennas [316] and those with fractal geometries [317]. Many other shapes can be used. Basically, all "particles" used to fabricate metamaterials can also function as nanoantennas. This includes single and double split rings, crescent antennas, etc.

An important group of nanoantennas are those based on the Babinet principle. A metal shape surrounded by dielectric and a dielectric-filled hole in metal with identical shape and size have identical diffraction patterns. Thus bowtie shaped holes in metal can be used, two holes are a Babinet equivalent of a nanodimer, arrays of nanoholes are equivalent of arrays of nanoparticles, crossed arrays of nanoholes correspond to crossed arrays of nanoparticles, etc. [318].

Nanoantennas are used in photodetection to couple propagating and localized modes and to localize fields. Especially, strong localizations are those in the feed gaps between coupled nanostructures. Photodetector enhancement by nanoantennas is a hot topic and numerous works have been published on it [243]. Earliest publications date as far as 1970s [319].

One of the approaches to the use of nanoantennas in photodetection is to use a Schottky metal-semiconductor junction. An optical antenna forms the metal part of the metal-dielectric contact at the semiconductor detector surface [320]. Photoexcitation generates hot electron-hole pairs by plasmon decay and the electrons are injected over the Schottky barrier, thus directly generating photocurrent. A problem with this approach is its low efficiency.

2.17 Plasmonic Nanoparticles and Redshifting

As the simplest nanoantennas, plasmonic nanoparticles can be utilized to enhance the absorption within thin-film solar cells [243]. They couple incoming waves with the localized SPP field, have increased scattering cross-section and strongly localize electromagnetic field just in the thin active region of the detector, Fig. 2.62. The same principle is applicable for infrared detection [321]. This cannot be done with pure noble metal nanoparticles since their surface plasmon resonance is in ultraviolet or visible part of the spectrum. Because of that their response must be redshifted. In this part, two approaches to such redshifting are described.

As Fig. 2.62 shows, the scattering cross-section of a plasmonic nanoparticle is greatly enhanced by plasma resonance. It can readily reach an order of magnitude value larger than the geometrical cross-section. The scattering cross-section for plasmonic nanoparticles at a wavelength λ can be calculated as [322]

$$C_{\mathrm{scat}} = \frac{8}{3\pi}\alpha^2\left[\frac{\pi}{\lambda}\right]^4, \tag{2.56}$$

where

$$\alpha = 3V\frac{\frac{\varepsilon_{\mathrm{np}}}{\varepsilon_{\mathrm{d}}}-1}{\frac{\varepsilon_{\mathrm{np}}}{\varepsilon_{\mathrm{d}}}+2}. \tag{2.57}$$

where $\varepsilon_{\mathrm{np}}$ is the complex relative dielectric permittivity of the plasmonic nanoparticles, determined by the Drude formula, ε_{d} is the permittivity of the dielectric

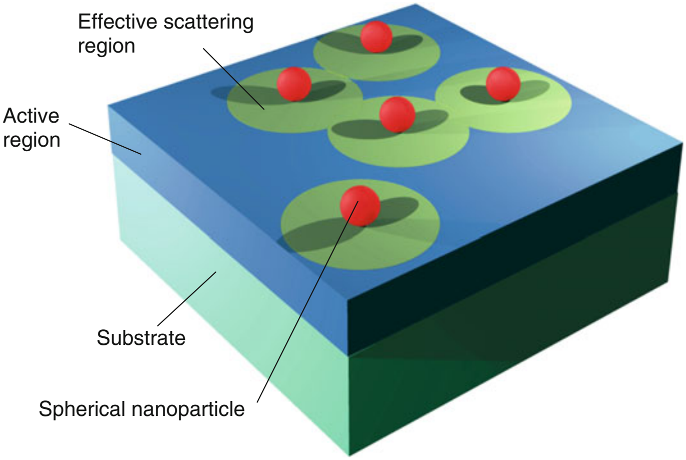


Fig. 2.62 Increase of scattering by nanoparticles on detector surface

medium, and V is the nanoparticle. The plasmon resonance and thus the maximum value of the scattering cross-section are reached at $\varepsilon_{np} = -2\ \varepsilon_d$.

The absorption cross-section is

$$C_{abs} = \frac{2\pi}{\lambda}\mathrm{Im}(\alpha). \tag{2.58}$$

Figure 2.63 shows several schemes for the enhancement of the performance of an IR detector by using nanoparticles and redshifting strategies. Figure 2.62a shows the case when alternative material (transparent conductive oxide like ZnO, ITO, and similar) [243] is used instead of metal nanoparticles. The plasma frequency of TCO nanoparticles is tunable by doping and by the fabrication procedure and in principle reaches longer wavelength (in infrared), while at the same time losses are smaller [243]. Another approach is to use alternative geometries of nanoparticles. Different metal-dielectric combinations and different geometries result in widely varying values of plasma frequency. For instance, nanoshells can have plasma frequency from visible to 9 μm, nanotriangles reach 5 μm, and nanorice 2 μm [323]. A core-shell nanoparticle is shown on the top of the active region (Fig. 2.62b). Instead of nanoparticles with alternative geometry and composition, one can use nanoparticles embedded in high-permittivity dielectric (requires additional antireflection layer at the surface), Fig. 2.62c. Plasmonic nanoparticles can be also placed at the back surface of the detector or within its active region, and not at the top of the detector, Fig. 2.1d. Further redshifting can be obtained by adjusting the distance between the nanoparticles. It has been shown that smaller interparticle distance results in a larger redshift [324]. Obviously, two or more of the quoted strategies can be used simultaneously.

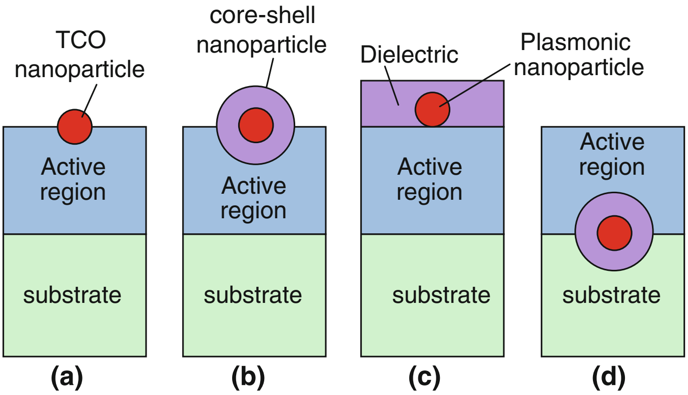


Fig. 2.63 Some schemes for infrared IR detector enhancement by plasmonic nanoparticles. **a** Use of transparent conductive oxide at the surface; **b** core-shell plasmonic nanoparticle embedded in high-permittivity dielectric at the surface; **c** high-permittivity dielectric layer with embedded nanoparticle; **d** core-shell plasmonic nanoparticle embedded in high-permittivity dielectric at the back surface

Fig. 2.64 Spectral dependence of scattering S_{scat} cross-section for an embedded ITO particle, r = 60 nm, λ_p = 625 nm, substrate permittivity 10

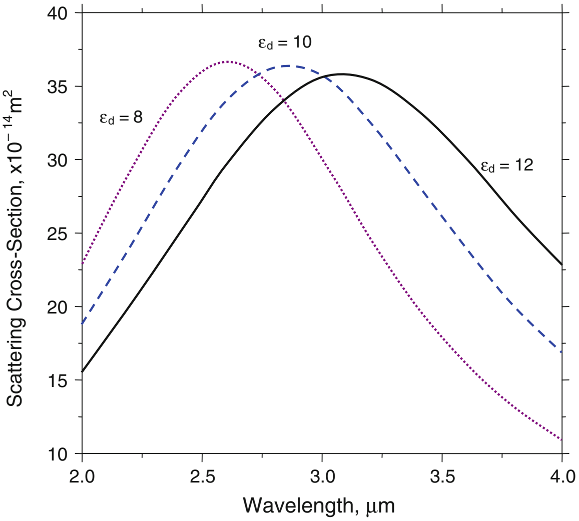


Figure 2.64 shows the calculated scattering cross-section for a spherical dipole indium tin oxide nanoparticle with a radius of 60 nm at a substrate with a relative dielectric permittivity of 10. The doping concentration was 1.2×10^{21} cm^{-3}, with an effective mass of m* = 0.4 m_0, thus plasma frequency was 4.8×10^{14} Hz. The nanoparticle is located at the top of the active surface of the detector and is embedded in dielectric (the values are shown near each respective curve). Finite

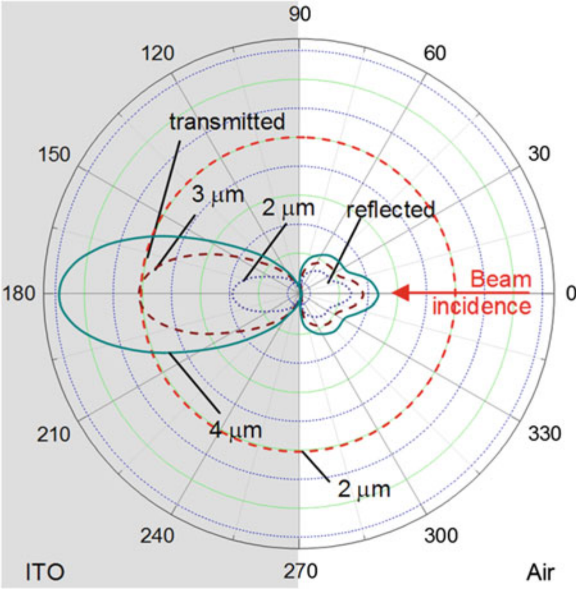


Fig. 2.65 Polar diagram of far field distribution of electric field due to scattering around a spherical ITO nanoparticle obtained by finite element method. $r = 60$ nm, $\lambda_p = 625$ nm, operating wavelengths 2, 3, and 4 μm. Permittivity of the dielectric layer is 10

element method was utilized for calculation. A redshift of the peak of the scattering cross-section increases with the permittivity of the embedding dielectric.

Figure 2.65 shows the radial distribution of the scattered electric field. Forward and back scattering are readily seen. It is visible that scattering increases with the operating wavelength. An increase of the permittivity of the embedding dielectric layer also causes an increase of scattering.

Nanoparticles can be ordered at the detector surface or elsewhere using self-assembly techniques or pattering by top-down approach. They can be distributed in regular patterns, thus accurately controlling the interparticle distance or they can be randomly scattered. A possible way to implement nanoparticles for plasmonic enhancement is to arrange them in a quasicrystal pattern (for instance Penrose tiling), which ensures an isotropic photonic response of the structure [325].

Chapter 3
Charge Carrier Management (Thermal Noise Engineering)

3.1 Nonequilibrium Suppression of Auger Generation-Recombination

3.1.1 Physical Principles

There is a single concept behind all of the existing nonequilibrium methods for IR detector performance improvement. A modification of charge carrier transport within a photodetector is done with the aim to cause a local equilibrium disturbance between the carriers and the crystal lattice. This is used to suppress minority carrier concentration below its equilibrium value. As a result, carrier concentration may reach values several orders of magnitude below the equilibrium one at near-room temperatures. As far as the carrier generation-recombination is concerned, the effects of nonequilibrium concentration decrease are equivalent to photodetector cooling.

Nonequilibrium suppression of generation-recombination occurs only in a relatively narrow range of carrier energies. If carrier temperature is equal to the lattice temperature (T_L) or close to it, the nonequilibrium is insufficient to cause significant suppression. If opposite is the case, hot carrier effects (e.g., impact ionization) may obliterate any benefits stemming from the nonequilibrium transport.

The use of nonequilibrium methods results in a decrease of majority carrier concentration to a level near the extrinsic one, while the concentration of minority carriers is additionally decreased to maintain electroneutrality. This results in suppression of all three main recombination processes. Auger generation is also decreased to a certain extent. Thus, the total g-r noise is suppressed and the device approaches the BLIP limit if its performance is g-r process-limited (which is, according to 1.6.2, the case in most of the practical situations). At the same time, cryogenic cooling is either completely absent or can be applied in a significantly lesser degree—the operating temperatures of nonequilibrium devices for IR detection are typically above 200 K.

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2_3

Auger processes are most strongly influenced by concentration decrease, since they are proportional to n^2p or p^2n. This is the reason why nonequilibrium methods are best suited for the devices with dominant Auger processes.

When applying nonequilibrium methods it is important not to increase the probability of impact ionization, i.e., the effective temperature of carriers must not be elevated too much. Nonequilibrium methods are used for relatively weak fields, out of the hot electron range, but near to it.

An important consequence of the use of nonequilibrium methods is that they enable the operation of bipolar and unipolar devices in narrow-bandgap semiconductors at elevated temperatures and offer the possibility to use doping level adjustment to control their performance. This is called pseudo-extrinsic behavior.

An advantage of the use of nonequilibrium methods for IR photodetectors is that the concentration decrease in the active area directly increases sensitivity, since now the concentration increase caused by irradiation amounts to a larger percent of the total concentration.

Carrier transport in nonequilibrium devices is modified and its nonequilibrium distribution achieved by joint action of internal and external fields. Any of the possible causes of carrier transport may be included, i.e., electric and magnetic field, temperature gradient, chemical potential gradient, and their combinations.

Contrary to the cases of cryogenic cooling of detectors and the application of optical methods for the increase of radiative lifetime, the device in operating mode is here in dynamical state. Although kinetic processes in detectors in a general case may be nonstationary (various transient phenomena and oscillatory modes of operation, e.g., in detectors utilizing the traveling wave mechanism [326] or hot electrons [327]), this work is dedicated solely to stationary phenomena.

All of the nonequilibrium methods presented in the literature until now used only electric field or a combination of electric and magnetic field to improve photodetector performance.

The idea of causing nonequilibrium in carrier distribution to decrease their concentration and/or cool a structure appears in the literature in different contexts. A number of papers handle the effects of recombination suppression in strong fields and at low temperatures. Nawrocki et al. reported significant suppression of Auger recombination and lifetime increase in semimagnetic semiconductor MnCdS in strong magnetic fields and at cryogenic temperatures [328]. Imamoglu and Yamamoto [329] studied shot noise suppression in p-i-n structures at mesoscopic level (i.e., with a thickness smaller than the nonelastic scattering length) and arrival to a sub-Poissonian distribution, and de Jong and Beenakker proposed a semiclassical theory of this noise suppression [330].

This work handles solely non-cryogenic temperatures, and the accent is given to room temperature operation. At that, transport phenomena are studied only at macroscopic level.

A partial analogy may be drawn between the electron characteristics of nonequilibrium detectors and semiconductor lasers. Actually, in some aspects of their statistics and carrier transport the nonequilibrium detectors may be described as "inverse lasers". To suppress carrier generation-recombination it is necessary to reach a state opposite to that of inverse population of energy bands. An ideal situation would be achieved if some sort of nonequilibrium "pumping" were used to completely deplete the conduction band, i.e., to reach a distribution corresponding to temperatures close to absolute zero. Both in lasers and in photodetectors it is necessary to use strong injection mode and large deviations from equilibrium. In both cases it is necessary to ensure the best possible confinement of carriers. A difference is that in the case of lasers our desire is to maximize the carrier concentration within the active region, while in detectors we need to minimize it.

It is only to be expected that some nonequilibrium detector structures have their analogs in semiconductor lasers. Exclusion detectors correspond to single-heterolasers, extraction devices to double-heterolasers, and magnetoconcentration detectors to lasers with the "magnetoelectric photoeffect" proposed by Marimoto et al. [331]. This inverse analogy is valid not only in electrical, but also in optical field, where e.g., resonant cavity (RCE) detector structures are connected with VCSEL lasers, and lasers with a PBG cavity with PCE (photonic crystal-enhanced) detectors.

It is necessary to fulfill two conditions to cause nonequilibrium with an aim to decrease carrier concentration and thus to decrease g-r noise. One of them is the existence of a mechanism to cause carrier transport (generally, external or internal fields or their combinations, resulting in a force acting to charge carriers). The second one is the existence of a layer (barrier) on the surface or in the bulk actively or passively removing carriers from the active region or preventing them to enter it. However, some mechanisms may simultaneously act in both mentioned ways.

Table 3.1 shows some possible methods of application of external and/or built-in fields for generation-recombination suppression in narrow-bandgap materials. As said before, we limit ourselves to photoconductive and photovoltaic structures. Even so, it is visible that there is a number of nonequilibrium structures and modes of operation never previously investigated.

It should be mentioned that Table 3.1 gives not only references directly connected with nonequilibrium photodetectors and noise suppression, but also with related phenomena, among which negative luminescence is the most important one. For instance, the use of thermal gradient was considered connected with InSb luminescence, but not with photodetector performance.

Table 3.1 Some transport mechanisms utilized for g-r processes suppression in photonic detectors based on narrow-bandgap semiconductors

Mechanism	Photoconductors	Photovoltaics
$\vec{E}$	Exclusion [332]	Extraction [333]
$\vec{E} \cdot \vec{B}$	Magnetoconcentration [334]	(*)
∇T	[335]	(*)
$\nabla T, \vec{E}$	(*)	(*)
$\nabla T, \vec{E} \cdot \vec{B}$	(*)	(*)
$\nabla T, \vec{B}$	(*)	(*)

Cases denoted with (*): no literature entries

Table 3.2 Basic blocking structures used in nonequilibrium IR photonic detectors

Blocking structure	Type	References
$n^+\nu$ or $p^+\pi$ structure (l-h junction)	Exclusion contact	[332]
$N\nu$ or $P\pi$ structure (single-heterojunction)		
$n^+\nu p^+$ or $n^+\pi p^+$ structure	Extraction + exclusion	[336]
$N^+\nu P$ or $P^+\pi P$ structure (double-heterojunction)	Contact + possibly	
	Barrier junction(s)	
S_{high}/S_{low} surfaces	Enhanced surface recombination	[334, 337]

Some structures used for blocking carrier transport to the excluded/extracted region are given in Table 3.2. Only generic types are given, and in some cases a number of different detectors stem from a single generic type.

3.1.2 Some General Properties of Nonequilibrium Photodetectors

Nonequilibrium detectors share a number of similar or even identical characteristics, independently on a particular structure or process. This is a consequence of similar basic principles utilized for nonequilibrium operation. All of the Auger-suppressed devices utilize relatively strong external or internal fields to decrease minority carrier concentration in a given volume and thus operate in the mode of large deviations from equilibrium. The degree of suppression is proportional to the intensity of the applied fields. The structures with the decreased concentration of minority carriers are conventional intrinsic photonic detector structures.

In all nonequilibrium devices presented until now the active region is fabricated in weakly doped ν or π material. If a nonequilibrium mechanism is applied the majority concentration in this region decreases near to extrinsic concentration. To maintain electroneutrality, the minority carrier concentration drops several orders of magnitude more. Thus, a nonequilibrium and stationary carrier distribution is reached and dynamically maintained by means of external fields. In such a mode semiconductor behaves again as an extrinsic one. This means that Auger-suppressed devices operate in nonequilibrium mode.

A decrease of carrier concentration in the active region increases the dynamic resistance of the detector. This is valid both for photoconductors and photovoltaics. For larger deviations of carrier concentration from the equilibrium value the dynamic resistance becomes also larger, so that its increase represents a figure of merit of a particular nonequilibrium method. If the number of carriers within the active region is low enough, the differential resistance changes its sign and becomes negative. With a further increase of bias the resistance changes its sign again and the current through the structure rises. The optimum operating point of a nonequilibrium device is in the inflection point of differential resistance, i.e., for the minimum dark current and thus the minimum noise current.

The next phenomenon connected with nonequilibrium methods is negative luminescence [337], also called induced absorption [338]. Actually, a semiconductor with nonequilibrium decrease of carrier concentration emits less electromagnetic radiation than it should according to the Planck law.

Finally, there is a phenomenon observed in practically all cases of nonequilibrium devices reported in the literature and limiting their application. It is the appearance of excessive 1/f (flicker) noise. This noise is proportional to the current through detector, and all the nonequilibrium methods utilized until now, independently on a particular mechanism, required increased current densities to reach nonequilibrium mode of operation. It is maintained that 1/f noise is a consequence of current flow through barriers at the surface or in the bulk of semiconductor and at Ohmic contacts and that convenient technological procedures could reduce it. However, in this moment it still appears to be a serious problem of the operation of nonequilibrium devices.

3.2 Limits of Nonequilibrium Detector Operation

In this section, we consider the limits of operating parameters required for the function of a nonequilibrium detector. This is a general consideration in the sense that it does not depend on a particular mechanism of nonequilibrium suppression.

A nonequilibrium device is limited from one side with the requirement to deplete carriers from it down to a level useful for generation-recombination noise suppression. On the other hand, this process requires relatively high internal and external fields and large deviations of carrier distribution from equilibrium. Thus, the applied fields themselves pose the upper limit of operation, mostly by causing carrier heating and increasing the probability of impact ionization to a point where it removes the benefit of Auger suppression. For a nonequilibrium device to function at all, a working range must exist where both of the above requirements are simultaneously fulfilled.

3.2.1 Lower Limit of Operation of Nonequilibrium Devices

The minimum of necessary carrier concentration decrease in a nonequilibrium device may be estimated utilizing the g-r processes theory presented in Sect. 1.4. We assume that SR processes are sufficiently decreased by convenient technological procedures, since this is at the same time the general condition for the applicability of nonequilibrium methods. The level to which it is possible to decrease radiative generation rate is $n_i/2\tau_{ri}$. Thus, there is no sense in decreasing the sum of modules of Auger generation and recombination rates significantly below this value. Therefore, we introduce the ratio between τ_{Ai} and τ_{ri} as a figure of merit of nonequilibrium detectors of arbitrary type. We call this parameteras the nonequilibrium depletion factor.

For instance, to bring an n-type detector to the limit of Auger process suppression, electron concentration must be either equal to the product of nonequilibrium depletion factor and intrinsic concentration or lower than it (if the Shockley-Read noise component can be neglected). Thus, the product between the nonequilibrium depletion factor and intrinsic concentration represents the highest allowed level of doping of a given semiconductor for a given temperature.

Figure 3.1 shows the dependence of the nonequilibrium depletion factor for an n-type device on cutoff wavelength, where intrinsic Auger lifetime was calculated according to the Beattie-Landsberg expression (1.51), and the integral determining intrinsic radiative lifetime was numerically calculated according to (1.37).

The calculated values in Fig. 3.1 are given for HgCdTe samples with their composition so adjusted to furnish a desired cutoff wavelength (required bandgap) at the operating temperature. It can be seen that, depending on the exact position of the absorption edge, it may be necessary to decrease concentration even three orders of magnitude. The conditions are more severe if the operating temperature is higher and the cutoff wavelength longer.

The value of nonequilibrium depletion factor equal to one means that for the corresponding cutoff wavelength (and all shorter values) it is not necessary to suppress g-r processes to achieve the BLIP specific detectivity. Figure 3.1 shows that at 77 K values larger than one are obtained in the whole range of interest 3–14 μm. This means that a detector cooled by liquid nitrogen does not require additional g-r process suppression, as is well known. However, at wavelengths above 14.55 μm this curve too drops below one; thus it is necessary to cool detectors for very long wavelengths below the liquid nitrogen temperature in order to reach BLIP operation.

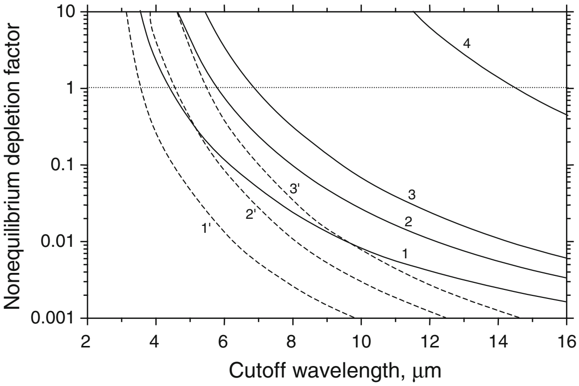


Fig. 3.1 Dependence of nonequilibrium depletion factor on cutoff wavelength for mercury cadmium telluride detector. Solid lines (numbers without prime): radiative lifetime calculated without reabsorption. Dashed lines (primed numbers): ninefold increase of lifetime due to reabsorption. Curve 1 300 K, curve 2 220 K, curve 3 180 K, curve 4 77 K

Fig. 3.2 Increase of radiative lifetime by PCE method (the ratio of the values of τ_{ri} with PCE and without it) on BLIP temperature and wavelength

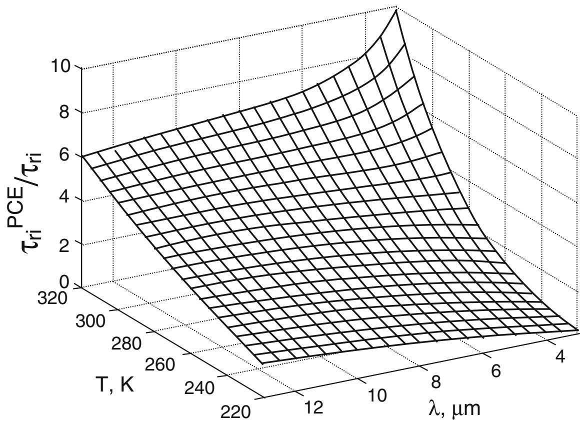


Figure 3.1 also shows the influence of reabsorption to improvement of nonequilibrium device characteristics. The increase of radiative lifetime is ninefold, which is achievable by structures with photonic bandgap. According to Sect. 2.14, the quoted improvement is readily achievable even by simple 1D structures. It can be seen that the conditions posed to the doping of such a detector are much more severe, which is a result of the BLIP limit being shifted toward larger detectivities.

Next we consider the bottom limit of the operation of uncooled BLIP in an alternative manner by analyzing the BLIP temperature. The following procedure was used to calculate this temperature: the radiative generation rate was made equal to the background generation rate, $n_i/2\tau_{ri} + \eta\Phi_B/d$ (where η is quantum efficiency, Φ_B is the background flux, typically given by Planck's law, d is the active area thickness). Figure 3.2 shows the dependence of the factor of PCE increase of radiative lifetime (the ratio of τ_{ri} with photonic crystal enhancement and without it) on BLIP temperature and wavelength.

The required increase factor at lower temperature becomes larger with wavelength, to reach a point of inflection at temperatures near the room temperature. Subsequently for larger values of λ this factor decreases with wavelength.

Elliott et al. [339] analyzed noise mechanisms in MWIR and LWIR infrared detectors operating in the range from 3–13 μm and proved that there is no fundamental obstacle that would prevent room temperature operation of photodetectors in background limited performance, even if the field of view is reduced.

3.2.2 Maximum Fields for Nonequilibrium Suppression

The upper limit of operation of nonequilibrium devices is set by the effects of hot carriers—naturally, under the assumption that the effects of Joule heating is controlled by proper heat sinks and by structures with high-thermal conductivity.

Since carrier heating caused by strong fields is antagonistic to the nonequilibrium depletion of the active region, and both processes are proportional to applied fields, in some point their action will become equal and remove the benefits of Auger processes. Besides that, since the main mechanism of nonradiative carrier generation is impact ionization, there is a danger of irreversible breakdown and device destruction.

We will assume that the temperature in the expressions for Auger generation and recombination (as given in Sects. 1.4.2 and 1.4.6) actually represents carrier temperature. The increase of external fields above the limit of hot carrier effects means an increase of Auger generation. At the same time expressions for it as quoted in Sect. 1.4 cease to be valid.

The influence of applied electric field to carrier temperature increase will depend on scattering processes defining the transport parameters. Carriers are not heated as long as the quasi-free charge carriers are in equilibrium with crystal lattice. This means that during a collision they transfer all their energy to the lattice, i.e., the energy acquired by an electron on its free path must be smaller or equal to the maximum energy transferable to the lattice in a single collision.

The kinetic energy that can be transferred to the lattice in a single collision can be represented as $\delta K = \delta P^2/2\ m^* = (Ft)^2/2\ m^*$. In a purely electric field the force F is equal to qE. The time t represents the period of crossing of free path, i.e., the mean relaxation time in scattering process, $\langle\tau\rangle$. This means that $\langle\delta E^*\rangle = (q\langle\tau\rangle E)^2/2\ \mathrm{m}^*$. The relaxation time $\langle\tau\rangle = \mu \mathrm{m}^*/\mathrm{q}$ is a complex function depending on several different scattering mechanisms. The mobilities of HgCdTe and InSb are large, as well as in other narrow-bandgap materials, meaning that electrons in them readily reach the hot carrier region.

One of the approaches to estimate the effects of scattering mechanisms is to consider carrier temperature increase due to a particular mechanism [340]. In narrow-bandgap materials for a temperature range between 200 K and room temperature the prevailing mechanism is polar optical phonon scattering [13]. The increase of carrier temperature due to this mechanism is [341]

$$\frac{\Delta T}{T} = \frac{2m_n^* \mu_n^2 E^2 K_1\left(\frac{T_D}{2T}\right)}{3kT_D K_0\left(\frac{T_D}{2T}\right)}, \tag{3.1}$$

where T_D is the Debye temperature for the given semiconductor material, while K_0 and K_1 denote the modified Bessel functions of the second kind of the zeroth and first order.

The influence of electric field to the increase of Auger generation can be calculated starting from the Beattie-Landsberg expression for Auger 1 intrinsic recombination lifetime τ_{A1}^i

$$\tau_{A1}^i = \varphi\left(m_n^*, m_p^*, \varepsilon_s\right)\left(\frac{E_g}{k_b T}\right)^{3/2} e^{E_g/k_b T}, \tag{3.2}$$

where we use φ to denote the part of the expression whose form is not necessary for our derivation. Since Auger 1 generation rate is given as

$$g_{A1} = \frac{n}{2\tau_{A1}^{i}}, \tag{3.3}$$

we may write

$$g_{A1} = F\left(m_n^*, m_p^*, \varepsilon_s\right)\left(\frac{k_b T}{E_g}\right)^{3/2} e^{-E_g/k_b T}, \tag{3.4}$$

where $F = \varphi^{-1} n/2$. We assume that effective masses of carriers m_n^*, m_p^*, bandgap E_g, electron concentration n and relative dielectric permittivity ε_s do not depend on carrier temperature, but only on the T_L.

If we differentiate the Auger generation rate over electron temperature, we obtain

$$\frac{dg_n}{dT} = \frac{Fe^{-E_g/k_b T}}{\sqrt{T}} + \frac{3Fk_b e^{-E_g/k_b T}\sqrt{T}}{2E_g}. \tag{3.5}$$

The ratio of (3.4) and (3.5) is

$$\frac{1}{g}\frac{\mathrm{d}g_n}{\mathrm{d}T} = \frac{E_g}{K_b T^2} + \frac{3}{2T} \tag{3.6}$$

If we multiply this expression by $\Delta T/T = dT/T$ which is defined by the Seeger expression, we obtain the ratio between the increase of the impact ionization rate and the equilibrium electron generation rate

$$\frac{\Delta g}{g} = \frac{2m_n^* \mu_n^2 E^2 K_1\left(\frac{T_D}{2T}\right)}{3kT_D K_0\left(\frac{T_D}{2T}\right)}\left(\frac{3}{2} + \frac{E_g}{kT}\right), \tag{3.7}$$

which is equivalent to the one given by Ashley [333]

$$\frac{\Delta g}{g} = \frac{2m_n^* \mu_n^2 E^2 K_1\left(\frac{T_D}{2T}\right)}{3kT_D K_0\left(\frac{T_D}{2T}\right)}\left(\frac{3}{2} + \frac{E_g}{kT}\right). \tag{3.8}$$

It can be written as

$$\frac{\Delta g}{g} = \left(\frac{E}{E_{\mathrm{krit}}}\right)^2. \tag{3.9}$$

For $Hg_{1-x}Cd_xTe$ we utilize the expression for electron mobility

$$\mu_n = 1.09 \times 10^5\, T^{-1.68} \tag{3.10}$$

The electron effective mass and the bandgap

$$m_n^* = m_0 \left[1 + 2m_1^2 \frac{q}{3} \left(\frac{2\pi}{h} \right)^2 \left(\frac{2}{E_g} + \frac{1}{E_g + \Delta} \right) \right]^{-1}, \tag{3.11}$$

$$E_g = 0.235 - 3.4 \times 10^{-4}\, T^2/(210 + T), \tag{3.12}$$

where Δ denotes the spin split-off energy (0.9 eV), P_1 is Kane matrix element $(9.0 \times 10^{-10}\mathrm{eVm}), m_0$ is the electron rest mass. The Debye temperature for $Hg_{1-x}Cd_xTe$ is 200 K.

Figure 3.3 shows the dependence of critical electric field in a mercury cadmium telluride detector (E_{krit} in expression (3.9)) versus cutoff wavelength for three different temperatures.

According to [333], it is acceptable for carriers to reach the "warm" region, i.e. the carrier temperature increase may amount maximally a half of the original temperature, $\Delta T/T \approx 1/2$. According to Fig. 3.3 and assuming that $\Delta g/g \approx 1/2$, we obtain for the limits of the allowed electric fields the values starting from a few hundreds V/cm at 200 K and ending on about thousand at 295 K (right-hand ordinate in Fig. 3.3).

The influence of magnetic field to the decrease of the allowed electric field in a detector due to carrier heating was calculated as follows. We assumed that instead of the force qE the Lorentz force $qE + q\mu_n EB$ acts on carriers (we assume a magnetic field perpendicular to electric. In that case, if (3.9) is still valid, E_{krit} is modified by a factor $(1 + \mu_n B)$.

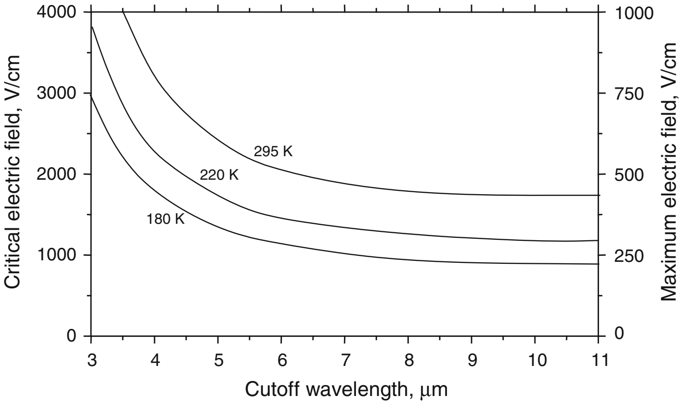


Fig. 3.3 Critical electric field in nonequilibrium HgCdTe detector versus cutoff wavelength at different operating temperatures

The temperature dependence of this factor for various values of magnetic induction and cutoff wavelength is shown in Fig. 3.4. For lower magnetic inductions this factor obviously tends to 1. It can be seen that magnetic field increase and temperature decrease significantly decrease allowed electric fields in detector. At $B = 2$ T and for $\lambda_{co} = 10.6$ μm we obtain that the limiting values of electric field at near-room temperatures are 2–4 times lower than without magnetic field, i.e. according to Fig. 3.4 the allowed electric field is of the order of 100 V/cm.

Applied fields act antagonistically to Auger process suppression through heat dissipation due to Joule effects of the carrier flow through semiconductor. Negative consequences of detector heating are numerous, and include parameter change including cutoff frequency shift, Auger generation increase and g-r noise increase, up to detector damage or even destruction.

Magnetic fields significantly add to the Joule heating increase. This arrives as a consequence of the inhomogeneity of carrier distribution within the sample caused by Lorentz force. The stronger magnetic fields, the larger part of carriers is forced to flow through narrower part of the sample and the result is a local increase of current density, proportional to magnetic induction.

Minimization of Joule effects is a serious problem in semiconductor microelectronic generally. Among the standard methods to reduce dissipation are the decree of the device volume, the improvement of the contact between active area and the substrate, mounting of a heat sink, optimization of the methods of detector mounting, etc. All of these methods can be utilized for nonequilibrium detectors. For instance, in the case of devices utilizing magnetic fields, one of the ways to reduce dissipation effects is to position a heat sink near the part of the sample with a local current density maximum.

To model performance change of nonequilibrium devices due to Joule losses an approximate empirical relation is applicable stating that the detector temperature increase is linearly proportional to the dissipation power divided by the active electrical area of the detector. For a HgCdTe photoconductor the proportionality factor has an empirical value of 0.6 Km^2/W). When numerically calculating thermal

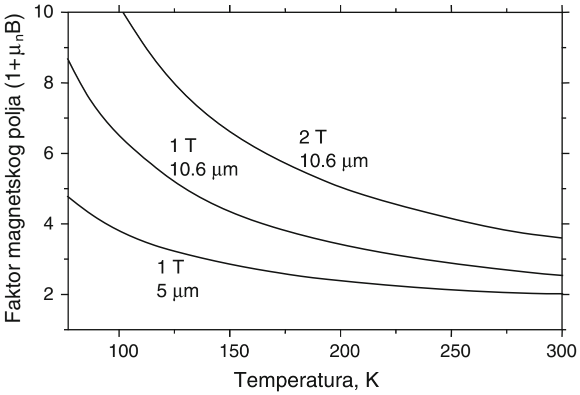


Fig. 3.4 Temperature dependence of magnetic field factor $(1 + \mu_n B)$ for $Hg_{1-x}Cd_xTe$

behavior of a nonequilibrium detector the temperature increase is added to the initial temperature of the photodetector, the temperature thus obtained is replaced into the model and thus a self-consistent procedure is formed used to iteratively calculate final detector parameters.

3.3 General Model of Nonequilibrium Photodetectors with Auger Suppression

Due to particularities of processes and materials used for Auger-suppressed infrared photodetectors, in their modeling it is not possible to straightforwardly apply conventional models based on small signal approximations and nondegenerate material properties. Therefore standard program packages utilized in simulation of semiconductor devices are not applicable. To models standard devices a set of fundamental equations is used, consisting of the Poisson equation, continuity equations, and current density equations, together with corresponding boundary conditions. When written for an equilibrium device utilizing solely electric field, this set of equations is known as the Van Roosbroeck model. To define the set of basic equations in the case of nonequilibrium photodetectors it is necessary to take into account a number of additional effects and thus to significantly modify the model.

All Auger-suppressed devices require for their function strong external fields (but they must not be so strong to cause hot-electron effects) and there is large deviation of their carrier concentration distribution from its equilibrium value. Instead of using conventional materials (Si), they typically use narrow-bandgap direct semiconductors with zincblende crystal lattice, which requires to modify the methods of calculation of their transport coefficients. Further, it is usually not possible to neglect carrier degeneration in nonequilibrium devices. Finally, in some types of nonequilibrium detectors it is necessary to take into account other fields besides electric (magnetic field).

The goal of the presentation in this section is to pose a generalized model comprising all the types of nonequilibrium detectors presented in the literature until now, and applicable to potential novel devices. When deriving the model we start from the semiconductor equations in their general form (Maxwell's equations and Boltzmann's transport equation.).

3.3.1 *Poisson Equation*

Poisson equation, i.e., the electroneutrality equation is basically the Gauss law, that is to say the Maxwell's equation giving the dependence between the electric induction vector and charge density

$$\operatorname{div} \vec{D} = \rho, \tag{3.13}$$

or

$$\operatorname{div}(\varepsilon \operatorname{grad} \psi) = -\rho, \tag{3.14}$$

where $\vec{D}$ is electric induction, ρ is charge density, ε is dielectric permittivity of semiconductor material, and ψ is electrostatic potential.

We assume that electric and magnetic field are quasistatic, i.e., that their changes are slow enough to assume that time derivatives of magnetic vector potential and scalar electric potential are approximately zero. This is not a too severe requirement, since according to [342] we may assume that a semiconductor device behaves quasistatically if the frequency of their fields are below several hundred of gigahertz.

Our further assumption is that the dielectric permittivity ε may be treated as scalar. This assumption is valid for all semiconductors with diamond and zinc-blende lattice [342, 343].

We further assume that the dielectric permittivity is spatially uniform. Our last assumption is that the magnetic permeability of the semiconductor material in question is approximately equal to the magnetic permeability of vacuum, i.e., that we use nonmagnetic material.

Spatial charge density is given by

$$\rho = q\left(p - n + N_D^+ - N_A^-\right), \tag{3.15}$$

where N_D^+ and N_A^- are the concentrations of ionized donors and acceptors.

3.3.2 Continuity Equation

The continuity equation is actually the Maxwell equation

$$\operatorname{rot} \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \tag{3.16}$$

where $\vec{H}$ is the magnetic field vector. By determining the divergence of the above expression and using (3.15) we obtain that $\operatorname{div} \vec{J} + \partial\rho/\partial t = 0$ (conservation of total charge). The total current density is split into the components of electron and hole current

$$\vec{J} = \vec{J}_p + \vec{J}_n. \tag{3.17}$$

We assume that the concentration of ionized dopants (the difference between the concentrations of ionized donors and acceptors) are not time dependent, i.e., we neglect “slow states.” This is not fully justified, since these states are actually responsible for flicker noise.

By introducing (3.17) into the divergence of (3.16) and simultaneously adding and subtracting net generation-recombination rate we obtain two continuity equations

$$\operatorname{div}\vec{J}_n - q\frac{\partial n}{\partial t} = -q\left(g_n - \frac{\delta n}{\tau_n}\right), \tag{3.18}$$

$$\operatorname{div}\vec{J}_p - q\frac{\partial p}{\partial t} = -q\left(g_p - \frac{\delta n}{\tau_n}\right), \tag{3.19}$$

where g is total generation-recombination rate. This term is actually the sum of the term of optical generation and the expressions (1.85) and (1.86) including Auger, radiative, and Shockley-Read processes.

3.3.3 Transport Equation (Current Density in Nonequilibrium Detectors)

When calculating carrier transport in a nonequilibrium device we start from the Liouville-Von Neumann expression for the density matrix [342] which reduces under certain conditions to the Boltzmann transport equation. These conditions reflect the semiclassical approach to the description of charge carrier transport and may be summed as [342]

1. Band theory of semiconductors is applicable and the theorem of effective mass is valid.
2. External fields slowly change compared to the electron wavelength, i.e., electron dynamics may be represented by classical physics.
3. We assume single-particle (Hartree-Fock) approximation [344].
4. We assume that collisions are instantaneous, i.e., that external fields do not change their momentum during a collision.
5. We neglect electron-electron interactions.
6. We assume that scattering probability is independent on external fields.

The Boltzmann equation is an integro-differential equation establishing a connection between disturbances in carrier distribution functions caused by external fields and scattering effects. Basically, this is a continuity equation in momentum space applied to carrier transport.

One of the ways to represent the Boltzmann equation is [345, 346]

$$\frac{\partial f}{\partial t} = -\nabla_{\vec{r}} f \vec{v} - \nabla_{\vec{k}} f \frac{\vec{F}}{\hbar} - \nabla_{\vec{k}} f \frac{\vec{F}_c}{\hbar} = \left(\frac{\partial f}{\partial t}\right)_{\text{difuz}} + \left(\frac{\partial f}{\partial t}\right)_{\text{spolj.sile}} + \left(\frac{\partial f}{\partial t}\right)_c, \tag{3.20}$$

where $\vec{F}$ denotes force caused by all external fields, $\vec{K}$ is momentum vector, $\vec{r}$ is radius vector. The nabla operators denote differentiation in r- and k-space, respectively. The index c denotes collisions. The first term from the left in the right side of equation describes the influence of diffusion to the distribution function f, the middle term the effects of external forces (a generalized drift term), and the third one (the right-most side) is the collision term. The latter may be represented as

$$\nabla_{\vec{k}} \frac{\vec{F}_c}{\hbar} = \left(\frac{\partial f}{\partial t}\right)_c = \int\limits_{k_f} \left[W\left(\vec{r}, \vec{k}_f, \vec{k}_i\right) f\left(\vec{r}, \vec{k}_f, t\right) - W\left(\vec{r}, \vec{k}_i, \vec{k}_f\right) f\left(\vec{r}, \vec{k}_i, t\right)\right] d^3k_f d^3k_i, \tag{3.21}$$

where W is probability of scattering of a particle in a unit time, while the index f denotes the final, and i the initial state.

The next assumption is that the relaxation time approximation is valid. According to Nag, this is equivalent to the assumption that scattering processes are elastic or isotropic and that quasi-Fermi levels can be used to describe carrier distribution [347]. In that case

$$\left(\frac{\partial f}{\partial t}\right)_c = -\frac{f - f_0}{\tau}, \tag{3.22}$$

where f_0 is quasi-equilibrium distribution function for the carrier type under consideration.

For a stationary case the Boltzmann transport equation in relaxation time approximation is

$$\frac{\vec{F}}{\hbar} \nabla_{\vec{k}} f\left(\vec{k}, \vec{r}\right) + \vec{v} \nabla_{\vec{r}} f\left(\vec{k}, \vec{r}\right) = -\frac{f\left(\vec{k}, \vec{r}\right) - f_0\left(\vec{k}, \vec{r}\right)}{\tau}, \tag{3.23}$$

where the first term from the left describes the influence of drift caused by external fields, while the second term describes the influence of diffusion, i.e., the dependence of transport on spatial gradient of carrier distribution.

Next we assume that the region under consideration is at least several free path lengths away from the device boundary [342] in order to enable the existence of a solution for f in the form of an infinite series. If the perturbation caused by the external fields is small, only the first term of the series can be kept, and thus the nonequilibrium distribution function becomes

$$f\left(\vec{k}\right) = f_0\left(\vec{k}\right) + f_1\left(\vec{k}\right), \tag{3.24}$$

where f_0 is the equilibrium part of the distribution function, described by the Fermi-Dirac distribution function.

By replacing (3.24) into (3.23) and using the definition of the Fermi-Dirac function we directly obtain for the particular terms

$$\nabla_{\vec{k}} f_0\left(\vec{k}, \vec{r}\right) = \frac{\partial f_0}{\partial E}\frac{\partial E}{\partial \vec{k}} = \frac{\partial f_0}{\partial E}\hbar\vec{v}, \tag{3.25}$$

$$\nabla_{\vec{r}} f_0\left(\vec{k}, \vec{r}\right) = \frac{\partial f_0}{\partial E}\frac{\partial E}{\partial \vec{r}} = \frac{\partial f_0}{\partial E}\left[\nabla E_f + \left(E - E_f\right)\frac{\nabla T}{T}\right], \tag{3.26}$$

$$\frac{\partial f_0}{\partial E} = -\frac{1}{k_b T}\frac{\exp\left(-\frac{E-E_f}{k_b T}\right)}{\left[1 + \exp\left(-\frac{E-E_f}{k_b T}\right)\right]^2}, \tag{3.27}$$

$$f_1\left(\vec{k}\right) = -\frac{\tau\left(\vec{k}\right)}{\hbar}\vec{F}\nabla_{\vec{k}}\left[f_0\left(\vec{k}\right) + f_1\left(\vec{k}\right)\right]. \tag{3.28}$$

The investigation of detectors with a temperature gradient surpasses the scope of this text. Therefore we assume isothermal conditions and neglect the term containing ∇T.

We further use the fact that the velocity of electrons with an isotropic effective mass is given as $\hbar\vec{k}\big/m^*$ and that external forces are caused only by electric and magnetic fields (Lorentz force), $\vec{F} = q\left(\vec{E} + \vec{v} \cdot \vec{B}\right)$.

A rough approximation for the f_1 is obtained if we replace the above definition of the external force and the development (3.25) into the expression for f_1 (3.28), and simultaneously neglect the gradient of f_1 obtained on the right side. Therefore $f_1 \approx -\partial f_0\big/\partial E\tau\left(\vec{k}\right)e\vec{E}\vec{v}$ We assume that the accurate expression has a similar form and thus we write the following form for the nonequilibrium part

$$f_1\left(\vec{k}\right) = -\frac{\partial f_0}{\partial E}\vec{X} \cdot \vec{v}. \tag{3.29}$$

The unknown vector $\vec{X}$ can be calculated according to the expressions (3.28) and (3.29) and the above obtained fact that in the first approximation the nonequilibrium distribution function does not depend on magnetic induction.

First, we determine the momentum gradient of the nonequilibrium part of the distribution function directly from the approximation for f_1 and (3.29) as $\nabla_{\vec{k}} f_1\left(\vec{k}\right) = -(\partial f_0/\partial E)\vec{X}\hbar/m^* - \vec{v}\partial\left(\vec{X}\partial f_0/\partial E\right)\big/\partial\vec{k}$. We determine the obtained value into (3.28), together with the above given definition for the velocity containing m^*, the expression for the Lorentz force, and the expressions (3.23) and (3.29). Further we utilize the assumption that the fields are weak enough to permit us to neglect the terms containing the product $\vec{E}\vec{X} \sim E^2$ We also neglect the term with $\left(\vec{v}\cdot\vec{B}\right)\vec{v}$. Finally, utilizing vector algebra rules we obtain

$$\vec{X} \approx \frac{\tau\left(q\vec{E}-\nabla E_f\right)+\frac{q\tau^2}{m^*}\left[\left(q\vec{E}-\nabla E_f\right)\times\vec{B}\right]+\frac{q^2\tau^3}{m^{*2}}\vec{B}\left[\left(q\vec{E}-\nabla E_f\right)\cdot\vec{B}\right]}{1+\frac{q^2\tau^2}{m^*}B^2} \tag{3.30}$$

The carrier concentration in the interval k, k + dk is by the definition

$$n_k = \frac{2}{(2\pi)^3} f\left(\vec{k}\right), \tag{3.31}$$

and the corresponding current density

$$d\vec{J}_k = qv\frac{\mathrm{d}V_k}{4\pi^3} f\left(\vec{k}\right), \tag{3.32}$$

so that the integral value of the current density for the whole momentum space is

$$\vec{J} = \frac{q}{4\pi^3}\int\limits_{V_k}\vec{v}f\left(\vec{k}\right)dV_k = \frac{q}{4\pi^3}\int\limits_{V_k}\vec{v}\left(-\frac{\partial f_0}{\partial E}\right)\left(\vec{X}\vec{v}\right)dV_k, \tag{3.33}$$

where we utilized the fact that the equilibrium part of the distribution function is even.

If we replace (3.30) into (3.33)we obtain the following expression

$$\begin{aligned}\vec{J} = q^2K_1\left[\vec{E}-\frac{1}{q}\nabla E_f\right]+\frac{q^3}{m^*}K_2\left[\left(\vec{E}-\frac{1}{q}\nabla E_f\right)\times\vec{B}\right]\\ +\frac{q^4}{m^{*2}}K_3\vec{B}\left[\left(\vec{E}-\frac{1}{q}\nabla E_f\right)\cdot\vec{B}\right],\end{aligned} \tag{3.34}$$

where K_1, K_2, and K_3 are kinetic coefficient dependent on the scattering mechanisms and the band structure of the utilized material. The last term on the right side becomes zero if the electric field vector and the concentration gradient are perpendicular to the magnetic field vector. We assume that this condition is valid for the rest of our consideration here.

We introduce here the notation $\sigma_B = q^2K_1$ and $\sigma_B\mu_H = q^3K_2/m^*$ · σ_B is effective conductivity, i.e., the semiconductor conductivity in the direction of electric field when there is an influence of magnetic field. In a general case its value depends on the intensity of magnetic induction vector. μ_H is Hall mobility, representing effective drift mobility under the influence of transversal electric field with an intensity equal to the intensity of the applied magnetic field. Its value is close to the value of the conventional drift mobility. Both parameters (σ_B and μ_H) are in general case anisotropic tensors.

If we introduce the above notation into (3.34) and eliminate the last term we obtain

$$\vec{J} = \tilde{\sigma}_B \left[\vec{E} - \frac{1}{q} \nabla E_f \right] + \tilde{\sigma}_B \tilde{\mu}_H \left[\left(\vec{E} - \frac{1}{q} \nabla E_f \right) \times \vec{B} \right], \tag{3.35}$$

where we use tilde to denote a tensor value. Further we assume that in narrow-bandgap semiconductors of InSb and HgCdTe type both of these coefficients are scalar, i.e., that there is no anisotropy of effective conductivity and Hall mobility along different crystallographic directions in these materials.

If magnetic field is equal to zero, we obtain directly from the above expression of the drift-diffusion equation for current density in a semiconductor device.

Further we consider the derivation of the expressions for the current density in the case when there are both electric and magnetic field. We assume that the electric field vector has an x-and a y-component. In order to be orthogonal, magnetic field can have only its z-component. Carrier concentration gradient may have x and y components, while in a general case the current density can have all three components.

We use the connection between the Fermi level and the carrier concentration [348]

$$\nabla E_{fn} = -q \frac{D_n}{\mu_n} \frac{\nabla_n}{n}, \qquad \nabla E_{fp} = q \frac{D_p}{\mu_p} \frac{\nabla_p}{p}. \tag{3.36}$$

We write further effective conductivity in magnetic field as [346]

$$\sigma_{Bn} = \frac{qn\mu_n}{1 + \mu_n^2 B^2}, \qquad \sigma_p = \frac{qp\mu_p}{1 + \mu_p^2 B^2}. \tag{3.37}$$

Utilizing the vector multiplication rules, rearranging the terms, and grouping the vector components along each Cartesian axis we obtain that the current density along the z-axis must be zero, so that the grouping of the remaining terms gives

$$\begin{aligned} \begin{bmatrix} J_{nx} \\ J_{ny} \end{bmatrix} = & \frac{q}{1 + \mu_{Hn}^2} \left\{ n\mu_n \begin{bmatrix} E_x \\ E_y \end{bmatrix} + n\mu_n \mu_{Hn} B_z \begin{bmatrix} E_y \\ -E_x \end{bmatrix} \right. \\ & \left. + D_n \begin{bmatrix} \partial n / \partial t \\ \partial n / \partial y \end{bmatrix} + \mu_{Hn} D_n B_z \begin{bmatrix} \partial n / \partial y \\ -\partial n / \partial x \end{bmatrix} \right\} \end{aligned} \tag{3.38}$$

By multiplying the scalar elements by a unit matrix, by further multiplying the obtained matrices with electric field vector and concentration gradient and by finally summing the terms with corresponding vector components we obtain

$$\begin{bmatrix} J_{nx} \\ J_{ny} \end{bmatrix} = \frac{q}{1+\mu_{Hn}^2}\left\{ \begin{bmatrix} n\mu_n & n\mu_n\mu_{Hn}B_z \\ -n\mu_n\mu_{Hn}B_z & n\mu_n \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} + \begin{bmatrix} D_n & \mu_{Hn}D_nB_z \\ -\mu_{Hn}D_nB_z & D_n \end{bmatrix} \begin{bmatrix} \partial n/\partial t \\ \partial n/\partial y \end{bmatrix} \right\}, \tag{3.39}$$

or, written in a compact form

$$\vec{J}_n = q\widetilde{\mu}_n\vec{E} + q\widetilde{D}_n\nabla_n. \tag{3.40}$$

This is simultaneously a form of the current density for the case of negligible magnetic field. We conclude that the influence of magnetic field to a narrow-bandgap semiconductor can be represented by introducing new, magnetic field-dependent anisotropic (tensorial) drift mobilities and diffusion coefficients.

The expression for the hole current density is

$$\vec{J}_p = \widetilde{\sigma}_{BP}\left[\vec{E} - \frac{D_p}{p\mu_p}\right] + \widetilde{\sigma}_{Bp}\widetilde{\mu}_{Hp}\left[\left(\vec{E} - \frac{D_p}{p\mu_p}\nabla_p\right) \times \vec{B}\right]. \tag{3.41}$$

An equivalent procedure for electron current density furnishes

$$\begin{bmatrix} J_{px} \\ J_{py} \end{bmatrix} = \frac{q}{1+\mu_{Hp}^2B^2}\left\{ \begin{bmatrix} p\mu_p & p\mu_p\mu_{Hp}B_z \\ -p\mu_p\mu_{Hp}B_z & p\mu_p \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} + \begin{bmatrix} D_p & \mu_{Hp}D_pB_z \\ -\mu_{Hp}D_pB_z & D_p \end{bmatrix} \begin{bmatrix} \partial p/\partial x \\ \partial p/\partial y \end{bmatrix} \right\}. \tag{3.42}$$

The above expressions for the hole and electron current densities correspond to the notation utilized in [349] and are convenient for numerical simulation of 2D structures.

Further we determine the separate components of the electron and hole current density vectors. It is necessary to eliminate the components of electric field and concentration gradient along the y axis from the expression for the current density along the x-axis. The same is valid in the opposite direction.

For the sake of brevity we introduce again the notation σ_{Bn} and σ_{Bp}. We can write then

$$J_{nx} = \sigma_{Bn}E_x + \frac{\sigma_{Bn}D_n}{n\mu_n}\frac{\partial n}{\partial x} + \sigma_{Bn}\mu_{Hn}E_yB_z + \sigma_{Bn}\mu_{Hn}\frac{D_n}{n\mu_n}B_z\frac{\partial n}{\partial y}, \tag{3.43}$$

$$J_{ny} = \sigma_{Bn}E_y + \frac{\sigma_{Bn}D_n}{n\mu_n}\frac{\partial n}{\partial y} + \sigma_{Bn}\mu_{Hn}E_xB_z + \sigma_{Bn}\mu_{Hn}\frac{D_n}{n\mu_n}B_z\frac{\partial n}{\partial x}, \tag{3.44}$$

$$J_{px} = \sigma_{Bp} E_x - \frac{\sigma_{Bp} D_p}{p\mu_p} \frac{\partial p}{\partial x} + \sigma_{Bp} \mu_{Hp} E_y B_z + \sigma_{Bp} \mu_{Hp} \frac{D_p}{p\mu_p} B_z \frac{\partial p}{\partial y}, \tag{3.45}$$

$$J_{py} = \sigma_{Bp} E_y - \frac{\sigma_{Bp} D_p}{p\mu_p} \frac{\partial p}{\partial y} + \sigma_{Bp} \mu_{Hp} E_x B_z + \sigma_{Bp} \mu_{Hp} \frac{D_p}{p\mu_p} B_z \frac{\partial p}{\partial x}, \tag{3.46}$$

Let us consider separately the electron and the hole current. First, we eliminate the x-component of the hole concentration gradient from the expression for the y-component of the hole current density. In this manner we obtain

$$J_{py} = \left(1 + \mu_{Hp}^2 B_z^2\right) \sigma_p E_y - \left(1 + \mu_{Hp}^2 B_z^2\right) \frac{\sigma_p D_p}{p\mu_p} \frac{\partial p}{\partial y} - \mu_{Hp} B_z J_{px}. \tag{3.47}$$

In an equivalent manner we utilize J_{nx} to eliminate the x-component of the electron concentration gradient from J_{ny}

$$J_{ny} = \left(1 + \mu_{Hn}^2 B_z^2\right) \sigma_n E_y - \left(1 + \mu_{Hn}^2 B_z^2\right) \frac{\sigma_n D_n}{n\mu_n} \frac{\partial n}{\partial y} - \mu_{Hn} B_z J_{nx}. \tag{3.48}$$

An inverse procedure furnishes

$$J_{px} = \left(1 + \mu_{Hp}^2 B_z^2\right) \sigma_p E_x - \left(1 + \mu_{Hp}^2 B_z^2\right) \frac{\sigma_p D_p}{p\mu_p} \frac{\partial p}{\partial x} - \mu_{Hp} B_z J_{py}, \tag{3.49}$$

$$J_{nx} = \left(1 + \mu_{Hn}^2 B_z^2\right) \sigma_n E_x - \left(1 + \mu_{Hn}^2 B_z^2\right) \frac{\sigma_n D_n}{n\mu_n} \frac{\partial n}{\partial x} - \mu_{Hn} B_z J_{ny}. \tag{3.50}$$

The next part of this derivation regards the special case of a sample with a constant concentration along the x-axis. This situation includes numerous practical situations and actually corresponds to all devices with one-dimensional dopant distribution. If there is no concentration gradient in the x-direction (there is neither a dopant concentration gradient nor field intensity changes along that direction) then we can neglect the derivative of the carrier concentration along the x-direction in the expressions for J_{px} and J_{nx}. Now we additionally introduce into these expressions the definitions of σ_n, σ_p and assume that $\mu_{Hn} = r_H \mu_n \approx \mu_n$ and $\mu_{Hp} = r_H \mu_p \approx \mu_p$, i.e., that the Hall mobility is approximately equal to the drift mobility. According to [350] the Hall constant for near-intrinsic InSb is in the range 1.01 to 1.08, and its value is also comparable in HgCdTe [351]. In this manner we directly obtain the electron current density in y-direction as

$$J_{px} = q\mu_p p E_x + \mu_p B J_{py}, \tag{3.51}$$

$$J_{py} = q\mu_p p E_y + \mu_p B J_{px} - q D_p \frac{dp}{dy}, \tag{3.52}$$

$$J_{nx} = q\mu_n n E_x + \mu_n B J_{ny}, \tag{3.53}$$

$$J_{ny} = q\mu_n n E_y + \mu_n B J_{nx} - qD_n \frac{dn}{dy}, \tag{3.54}$$

These are the expressions obtained by Lile in his phenomenological consideration of InSb under the influence of Lorentz force [58, 352].

In the direction of the Lorentz force the total current flow must be equal to zero

$$J_{ny} + J_{py} = 0 \tag{3.55}$$

so that $J_{ny} = J_y$

$$\begin{aligned} J_y = {} & \frac{q\mu_n\mu_p\left(\mu_n + \mu_p\right)np}{\mu_n n\left[1 + \left(\mu_p B\right)^2\right] + \mu_p p\left[1 + (\mu_n B)^2\right]} BE_x \\ & + \frac{\mu_n D_p n \frac{dp}{dy} + \mu_p D_n p \frac{dn}{dy}}{\mu_n n\left[1 + \left(\mu_p B\right)^2\right] + \mu_p p\left[1 + (\mu_n B)^2\right]} \end{aligned} \tag{3.56}$$

This is the final expression for the electron current density in a semiconductor sample under the influence of crossed electric and magnetic fields in the direction normal to both applied fields. In an absence of magnetic field it reduces to the expression for the total diffusion current across the semiconductor sample depth (there is no drift in the direction normal to the applied electric field).

3.3.4 One-Dimensional Model of Nonequilibrium Photodetector

According to the previous three sections we can write the model of a generalized nonequilibrium photodetector. Its form corresponds to that of the standard Van Roosbroeck-Shockley set of semiconductor equations [343], but with taking into account the influence of magnetic field.

For electrons in a semiconductor sample the time-dependent one-dimensional continuity equation is

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial y} + g, \tag{3.57}$$

with the net generation-recombination term given by (1.82 and 1.83).

Electron current density for the case when there is only electric field is

$$J_n = q\mu_n nE + qD_n \frac{dn}{dy}, \tag{3.58}$$

and when there are orthogonal electric and magnetic fields

$$\begin{aligned} J_y = & \frac{q\mu_n\mu_p(\mu_n + \mu_p)np}{\mu_n n\left[1 + (\mu_p B)^2\right] + \mu_p p\left[1 + (\mu_n B)^2\right]} BE_x \\ & + \frac{q\mu_n D_p n\frac{dp}{dy} + q\mu_p D_n p\frac{dn}{dy}}{\mu_n n\left[1 + (\mu_p B)^2\right] + \mu_p p\left[1 + (\mu_n B)^2\right]}, \end{aligned} \tag{3.59}$$

where it should be noted that the above expression is written in the direction of the Lorentz force, i.e., orthogonally to the direction of bias.

The equations for hole current density and the corresponding continuity equation are obtained in the equivalent manner.

It should be born in mind that the drift and Hall carrier mobility, as well as the diffusion constant, are dependent on the concrete scattering mechanisms in a given semiconductor and thus in a general case on the intensity of the applied electric and magnetic fields. Since the electric field generally depends on the position within the sample, this means that the mobility and the diffusion constant are the functions of the spatial coordinate.

It should also be noted that in degenerate semiconductor the diffusion coefficient and the drift mobility are connected by the modified Einstein relation as [345]

$$D_n = \frac{2E_{fn}}{3q}\mu_n. \tag{3.60}$$

Instead of the above expression one may use the general approach of Leonard and Michael [50] where the connection between these two parameters is made using the Fermi integral normalized using the intrinsic concentration to determine a correction for the Einstein relation as a derivative of the quasi-Fermi level over a logarithm of electron concentration

$$D_n = \frac{\mu_n kT}{q}\frac{\mathrm{d}(E_{fn}/k_b T)}{\mathrm{d}(\ln n)}. \tag{3.61}$$

For nondegenerate electrons and for holes (the latter being always nondegenerate) the standard Einstein relation is valid

$$D_n = \frac{kt}{q}\mu_n, \quad D_p = \frac{kT}{q}\mu_p. \tag{3.62}$$

The last expression used in the generalized model of a nonequilibrium device is the electroneutrality equation. For the case of partially ionized donors and fully ionized acceptors its form is

$$n + N_A = p + N_D^+, \tag{3.63}$$

with an assumption that the concentration of ionized donors is approximately

$$\frac{N_D^+}{N_D} = 1 - f_{FD} \tag{3.64}$$

where f_{FD} is the Fermi-Dirac distribution function.

We conclude that the general model of a nonequilibrium detector consists of two continuity equations and two current density equations plus an electroneutrality equation. We showed how this model under certain assumptions furnishes Lile's phenomenological equations for semiconductor samples under the influence of the Lorentz force. For the structures where the changes of their parameters along the x-axis cannot be neglected, (3.39) and (3.42) must be used, and in the most general case (3.35).

In the above derivation we skipped all considerations of boundary conditions because these conditions depend on concrete structures and geometries. Instead we analyze them each in its corresponding place in the further text.

3.4 Contact Phenomena and Galvanic Suppression of Auger Processes

3.4.1 Introduction and Physical Concept

The simplest and chronologically the oldest approach to decrease carrier concentration in semiconductor photodetectors in order to suppress the Auger processes is to use solely electric fields. Inhomogeneous distribution of charge carriers, besides being a result of the application of external electric fields, may be caused by the existence of

1. contact structures,
2. material inhomogeneity (chemical potential gradient), and
3. recombination zones (bulk or surface).

The contact structures mentioned above are actually metal-semiconductor junctions. Ohmic contacts do not belong to this group, since they do not change the ratio of minority and majority carriers in the total density of the current flowing through them. The Schottky contacts, however, do belong here. They may be of barrier type (rectifying) or of anti-barrier type (not rectifying).

The second type of structures are those with a gradient of chemical potential. These are various kinds of semiconductor junctions (isotype or p-n, homo- or heterojunction, abrupt, or graded).

Recombination zones usually represent a combination of two zones (typically on the detector surface) with a high and a low recombination rate, acting so to cause inhomogeneity of charge carrier distribution.

The literature published until now mentions only the nonequilibrium detectors utilizing the mechanisms 2 and 3, although, analogously to equilibrium devices, larger response speeds (i.e., wider bandwidths) may be expected in nonequilibrium devices utilizing e.g. Schottky junctions.

This chapter and the next one are dedicated to nonequilibrium devices utilizing solely the structures mentioned under 2, while the structure utilizing high/low recombination rates are considered in subsequent text, in the context of galvanomagnetic methods.

The terms “anti-barrier” and “non-directing junction” are used in the further text both for isotype semiconductor homo- or heterojunctions and for the corresponding Schottky junctions, while the terms “barrier” and “directing junction” are used to denote p-n junctions (both homo- and heterojunctions) and Schottky diode junctions.

In all cases of contact structures the cause of nonequilibrium is the band bending caused by the contact difference of potentials, i.e., by the built-in electric field. The external electric field created by bias causes carrier drift within the bent bands region and a result is carrier nonequilibrium.

In the case of semiconductor-semiconductor junctions there is a number of different cases, including high-low, homo- or heterojunctions, e.g. n^+p, p^+n, P-n, N-p, n^+-π, p^+-ν, $P^+π$, etc. In all of these cases the semiconductor junction may be abrupt or graded. In the case of metal-semiconductor contacts the junction may be with or without an inversion layer.

In contact (junctions) structures with a large enough current flow one can encounter the following types of nonequilibrium

1. injection (excess minority carriers are injected through a rectifying contact)
2. extraction (minority carriers are drawn—“extracted” through a rectifying contact)
3. accumulation (minority carriers are amassed near a non-rectifying contact), and
4. exclusion (minority carriers are prevented in flowing through a non-rectifying contact).

To suppress Auger g-r processes those phenomena can be used that decrease carrier concentration in semiconductor, i.e., exclusion and extraction.

The properties of the above-mentioned nonequilibrium processes are summed in Table 3.3.

Sometimes the term “extraction” is used in the literature to denote the phenomenon described in this work as “exclusion” and vice versa [345].

Table 3.3 Properties of nonequilibrium mechanisms based on junction phenomena

1	2	3	4	5
Type of nonequilibrium	Contact type	Suitability for Auger suppression	External electric field in n-type semiconductor	External electric field in p-type semiconductor
Exclusion	Antibarrier	Yes	E < 0	E > 0
Accumulation	Antibarrier	No	E < 0	E > 0
Extraction	Barrier	Yes	E > 0	E < 0
Injection	Barrier	No	E > 0	E < 0

Figure 3.5 shows the distribution of excess carrier concentration near to barrier and antibarrier junctions of the semiconductor-semiconductor and metal-semiconductor type.

Some advantages of purely electrical methods for causing nonequilibrium in Auger-suppressed detectors are given below.

- Detector structure is simple because it does not require nonstandard parts and additional external components.
- There are no disturbances to surrounding circuitry generated by magnetic fields, temperature gradients, etc.
- The devices can be fully fabricated using conventional planar and epitaxial technologies.
- Device dimensions are minimal (complete detector may have dimensions in micrometers, which is desirable from the point of view of power dissipation). External parts are either unnecessary or the requirements for them are significantly reduced in comparison to cooled devices.

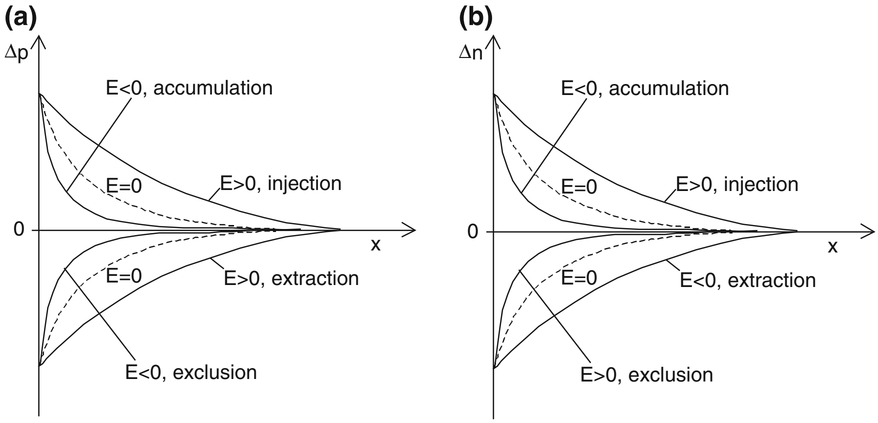


Fig. 3.5 Distribution of excess minority carriers in nonequilibrium junction structures. **a** Holes in n (or ν) type material; **b** electrons in p(π) region

3.4.2 Mathematical-Physical Model of Nonequilibrium Detectors Based on Minority Carrier Exclusion and/or Extraction: Boundary Conditions

In further text we consider a one-dimensional model with galvanic (i.e., purely electrical) Auger process suppression. Here, we actually continue the consideration presented in this chapter. Since the term of magnetic induction is here equal to zero, carrier transport is described by (3.57) and (3.58). Thus, the generalized model of a nonequilibrium detector is reduced to the van Roosbroeck's model [353]. This well-known model is the basis of programs for simulation of practically all standard semiconductor devices.

However, it should be mentioned that standard software packages cannot be used to model nonequilibrium devices. One of the two important reasons for that is the very principle of nonequilibrium devices operation, since by definition it requires large deviations from equilibrium and does not allow for standard approximations and linearizations while solving the model. The second reason is connected with the peculiarities of narrow-bandgap direct semiconductors used to fabricate infrared receivers for middle- and long-wavelength range. To implement semiconductor parameters properly into the algorithms for the solution of the models the effects of carrier degeneration must be taken into account, as well as the nonparabolicity of bands.

According to (3.57) and (3.58) we write the equations of the van Roosbroeck model for both carrier types, taking care that the parameters are dependent on a single spatial coordinate only (y).

(1) Continuity equation for electrons and holes

$$\frac{\partial n(y)}{\partial t} = \frac{1}{q}\frac{\partial J_n(y)}{\partial y} + g(y) \tag{3.65}$$

$$\frac{\partial p(y)}{\partial t} = \frac{1}{q}\frac{\partial J_p(y)}{\partial y} + g(y) \tag{3.66}$$

(2) Current density expressions

$$J_n(y) = q\mu_n(y)n(y)E(y) + qD_n(y)\frac{dn(y)}{dy} \tag{3.67}$$

$$J_p(y) = q\mu_p(y)p(y)E(y) + qD_p(y)\frac{dp(y)}{dy} \tag{3.68}$$

(3) Poisson equation

$$\frac{dE(y)}{dy} = \frac{q}{\varepsilon_0 \varepsilon_r}[n(y) - p(y) - (N_D(y) - N_A(y))] \tag{3.69}$$

The generation term for the case with zero optical excitation is calculated as a difference between (1.85) and (1.86). Optical generation is calculated using (1.80) or (1.81), while the influence of background is calculated using (1.82).

The boundary conditions for the galvanic-type Auger-suppressed devices are posed for the contacts in the points $y = 0$ and $y = l$. While deriving them we assume that the carrier distribution on contacts is always nondegenerate [343]. Let us denote with ψ_{ugr} the built-in diffusion potential, while U is bias voltage and ψ is the potential. If we apply the condition of equality of Fermi quasi-levels on contacts we obtain

$$\psi(l) - \psi(0) = U - \psi_{ugr}. \tag{3.70}$$

The next boundary condition is obtained based on the fact that the condition of quasi-neutrality is applicable to the contacts and that the carrier distribution near them is nondegenerate. For a situation when the semiconductor is of p-type in the point 0 the following is valid for minority carriers

$$n(0) = \frac{1}{2}\left(N_D(0) - N_A(0) + \sqrt{[N_D(0) - N_A(0)]^2 + 4n_i^2}\right) \tag{3.71}$$

$$p(0) = \frac{n_i^2}{n(0)} \tag{3.72}$$

If n-type semiconductor is used in the point d, the corresponding expressions are

$$p(l) = \frac{1}{2}\left(N_A(l) - N_D(l) + \sqrt{[N_D(l) - N_A(l)]^2 + 4n_i^2}\right) \tag{3.73}$$

$$n(l) = \frac{n_i^2}{p(l)} \tag{3.74}$$

An equivalent approach to the derivation of boundary conditions is also utilized in the case of isotype junctions, again using the corresponding minority carrier concentrations and expressions (3.71)–(3.72) or (3.73)–(3.74).

Further we consider the case when the surface recombination rates at the contacts are finite. In a special case of a barrier, i.e., p-n junction, we will denote these rates by S_n at the n contact and S_p at the p-contact. The boundary conditions (3.71)–(3.74) in this case transform into the new ones (denoted by ‘prime’)

$$p'(0) = p(0) - \frac{J_p(0)}{qS_n} = p(0) - \frac{p(0)n(0) - n_0p_0}{n_0 + p_0} \tag{3.75}$$

$$n'(0) = n(0) - \frac{J_p(0)}{qS_n} = n(0) - \frac{p(0)n(0) - n_0p_0}{n_0 + p_0} \tag{3.76}$$

$$p'(l) = p(l) - \frac{J_p(l)}{qS_p} = p(l) - \frac{n_0p_0 = p(l)n(l)}{p(l) + n(l)} \tag{3.77}$$

$$n'(l) = n(l) - \frac{J_p(l)}{qS_p} = n(l) - \frac{n_0p_0 = p(l)n(l)}{p(l) + n(l)} \tag{3.78}$$

Alternatively, the one-dimensional van Roosbroeck's model can be reduced to a single expression over one type of carriers. To this purpose we eliminate electric field and potential from the system. According to (3.65)–(3.68) for a stationary case the continuity equations become

$$\begin{gathered}\frac{\mathrm{d}}{\mathrm{d}y}\left(D_n\frac{\mathrm{d}n}{\mathrm{d}y}\right) + \frac{\mathrm{d}}{\mathrm{d}y}(nv_n) + g(y) = 0\\ \frac{\mathrm{d}}{\mathrm{d}y}\left(D_p\frac{dp}{dy}\right) - \frac{\mathrm{d}}{\mathrm{d}y}(pv_p) + g(y) = 0\end{gathered} \tag{3.79}$$

where $v_n = -\mu_n E$ is electron drift velocity, while $v_p = \mu_p E$ is hole velocity. Equations (3.79) and (3.80) include electric field and its first derivative via v_n and v_p. First derivative of electric field is obtained directly from the Poisson's equation (3.69). Electric field is eliminated by first summing the current density equations (3.67) and (3.68), which gives an expression for the total current density

$$J = (p\mu_p + n\mu_n)qE - D_p\frac{dp}{dx} + D_n\frac{dn}{dx} \tag{3.80}$$

From this expression E is calculated and then replaced into the continuity equation (3.79) or (3.80).

The total current density through the structure J is used as a new boundary condition instead of (3.75)–(3.78).

3.5 Exclusion Photoconductors

Exclusion of minority carriers is a nonequilibrium transport effect occurring as a consequence of the application of electric field to an isotype contact, i.e., to a junction between semiconductors of the same type, but with different dopant concentrations and/or bandgaps. In n-type semiconductors exclusion is provided by forming homojunctions of $n^{+}\nu$ type (ν denotes near-intrinsic n-doped material) or Nν heterojunctions (N denotes n-doped semiconductor with a wider bandgap). Equivalent structures of p-type semiconductors are $p^{+}\pi$ and Pπ junctions. An illustration of a typical exclusion structure may be seen e.g., in Fig. 3.6a.

The mechanism of minority carrier exclusion in an $n^{+}\nu$ structure can be described in the following manner. Semiconductor material in the n^{+} region is degenerated already at relatively low carrier concentrations because the density of states in the conduction band is low (for instance in HgCdTe electrons are degenerated already at concentrations below 10^{18} cm^{-3}). Degeneration suppresses Auger 1 processes in the highly doped region (see Eq. 1.62) and the hole diffusion length becomes very low (of the order of 0.1 μm). In addition to that, the hole concentration in the ν region must be larger than in the n^{+} region, meaning that the diffusion leakage of holes from the highly doped region into the low-doped one will be negligible. Further, in a situation when no bias is applied, the electric field in the n^{+} region will be low, since the resistance of this region must be low because of the high electron concentration. This means that no built-in field will exist to cause drift motion of holes toward the ν-region.

When the exclusion contact $n^{+}\nu$ is reversely biased, a drift field is generated that drives holes away from the junction into the bulk of the ν-region. At the same time the above described mechanism will not allow their replenishment from the n^{+} region. Therefore, the hole concentration will be decreased in the part of the exclusion device starting from the $n^{+}\nu$ junction and ending at a point within the ν-region whose coordinate for a given temperature and a given semiconductor material will depend on applied bias.

An isotype junction does not represent a barrier for the flow of majority carrier. However, in order to minimize the space charge the electron concentration will also drop, and its level will be defined by the dopant concentration. Since according to (1.38) the level of the Auger processes is proportional on carrier concentration, it will also drop.

Of course, an equivalent consideration is also valid for isotype junctions in p-type material.

3.5.1 History of the Utilization of Minority Carrier Exclusion

Minority carrier exclusion at isotype junctions of "high-low" type (high doping-low doping) is first mentioned in 1950s [354], exclusively in connection with metal-semiconductor contacts used to form Ohmic connections. Exclusion was a parasitic effect to be eliminated. Lade and Jordan noted and theoretically described the nonlinearity of the current-voltage characteristics of exclusion junction under high bias [355]. Most of the subsequent publications were dedicated to the analysis of various junction geometries and special cases, including abrupt and graded junctions.

Minority carrier exclusion was chronologically the first nonequilibrium effect to be described within the context of its use in infrared detectors for Auger processes suppression. In 1985 British researchers Ashley and Elliott published a paper [356] that introduced the concept of nonequilibrium suppression of Auger processes (including generation-recombination noise) at near-room temperatures. In that paper they proposed the use of the exclusion effect to that purpose. In the same year White published a detailed numerical and experimental analysis of exclusion devices [50]. In that paper he presented an approximate analytical model for determination of most important parameters of exclusion photoconductors.

Results connected with practical implementation of the devices and improvements of the theoretical model were published by Ashley, Elliott, and White [332, 357]. They also described a structure utilizing an opaque mask and lateral electrodes for signal readout to additionally eliminate noise sources. In 1990 a "vertical" exclusion photodetector with a self-filtering layer was proposed utilizing the Burstein-Moss effect in an $n^{+}\nu$ structure to improve quantum efficiency [358].

3.5.2 Various Implementations of Exclusion Photoconductors

Figure 3.6a shows the basic configuration of an exclusion photoconductor. A narrow layer of highly doped (or, alternatively, wide bandgap) semiconductor that may be either p-or n-type is positively biased. On the other side of the isotype junction is a region of near-intrinsic semiconductor (either π or ν-type). The junction between these regions may be abrupt or graded. Exclusion empties the carriers from the layer near the "high-low" contact. The excluded region is wider for higher values of bias.

Figure 3.6b shows a modification of this structure, a vertical exclusion photoconductor. In this case infrared radiation passes through an n^{+} layer that at the same time serves as a passive Burstein-Moss filter in a manner also proposed for conventional cooled InSb photodiodes [359]. The main advantage of this structure is a higher ratio of the active area to the area where surface recombination appears, so that the relative portion of noise due to surface states to the total noise of the device is significantly reduced.

Finally, Fig. 3.6c shows an exclusion structure with a mask for noise reduction. The opaque mask is deposited over the photodetector area where exclusion effects are less pronounced or completely absent. In this manner the largest contribution to photocurrent is from the area with largest carrier depletion. Voltage readout electrodes are placed in the part with maximum depletion, thus further reducing the influence of other parts of the detector. This configuration is used with electronic circuitry that further reduces noise.

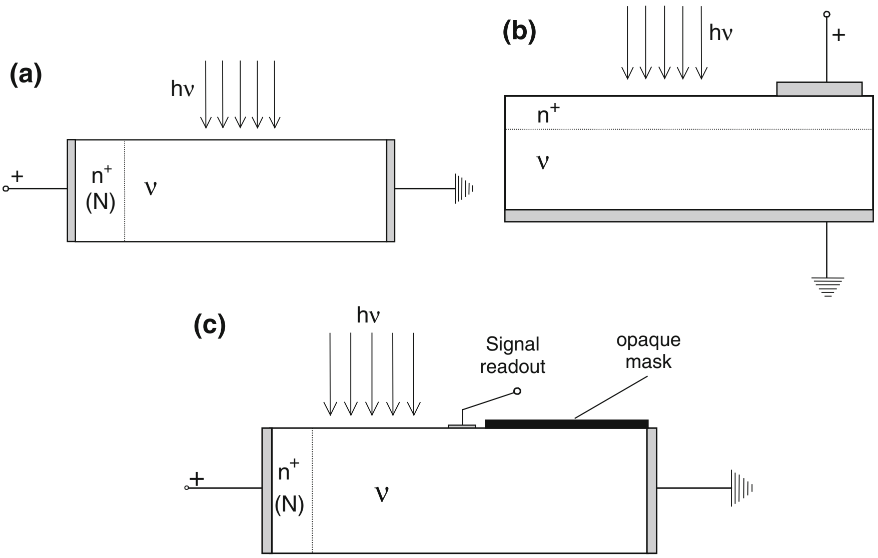


Fig. 3.6 Some implementations of photoconductors with minority carrier exclusion. **a** Basic configuration; **b** vertical structure with self-filtering and improved signal-to-noise ratio; **c** structure with opaque mask and lateral contacts

3.5.3 Approximate Analytic Calculation of Exclusion Detector Parameters

The further text presents an approximate consideration of transport phenomena within an exclusion device. Besides enabling us to simply determine the main parameters of exclusion structures, it also gives us a better insight in operation of Auger suppressed devices. It follows the approach of White [50], albeit with certain modifications that will later enable us to obtain some additional parameters.

First we analyze the $n^+\nu$ devices. Since the doping profile in the ν region is constant, the majority concentration ($\sim |N_D – N_A|$) and the minority concentration are also constant. The accuracy of this approximation is higher if the exclusion region is longer, because the influence of the transition region at the end of the ν-zone is proportionally smaller. The conclusion is that for a sufficiently depleted exclusion photodetector the generation-recombination rate must be approximately constant along the whole exclusion area, while the diffusion of both carrier types may be neglected. The continuity equation for majority carrier (3.79) is in that case

$$-\frac{\mathrm{d}}{\mathrm{d}y}(-n\mu_n E) + g = 0. \tag{3.81}$$

The expression for current density (3.58) directly gives the electric field intensity

$$E(y) = \frac{J}{q\left[p(y)\mu_p + n(y)\mu_n\right]}. \tag{3.82}$$

According to the continuity equation of minority carriers from (3.79), again neglecting the diffusion term, and using the conditions that electron and hole concentration gradients are equal and that $p = n - (N_D - N_A)$ we obtain for the majority carrier concentration gradient

$$\frac{dn}{dy} = \frac{qg}{J\mu_n\mu_p|N_D - N_A|}\left(p(y)\mu_p + n(y)\mu_n\right)^2. \tag{3.83}$$

It follows from the continuity equation (3.81) that

$$\frac{dE}{dy} = \frac{1}{n\mu_n}\left(-g - \mu_n E\frac{dn}{dy}\right). \tag{3.84}$$

Replacing the expression for the electric field (3.82) and the derivative of electron concentration (3.83) we obtain

$$\frac{dE}{dy} = -\frac{g}{\frac{\mu_p\mu_n}{\mu_n+\mu_p}|N_D - N_A|} \approx -\frac{g}{\mu_p|N_D - N_A|}, \tag{3.85}$$

$$E = E(0) - \frac{yg}{\mu_p|N_D - N_A|}, \tag{3.86}$$

where $E(0)$ is the electric field in a starting point $y = 0$. Therefore the electric field within the exclusion region is a linearly decreasing function only dependent on the doping level in the ν region, minority carrier mobility, and generation-recombination rate in that region.

Figure 3.7 shows the electric field distribution in an $n^+\nu$ nonequilibrium detector along the device with injected current density and dopant concentration as parameters. Equation (3.86) was used for the calculation. Recombination rate was calculated according to (1.85) and (1.86), assuming that the Auger and SR terms are negligible, which is consistent with the previous approximations.

Dashed lines show the values numerically calculated using the full model (3.67)–(3.69) and (3.79). It can be seen that the accurate results agree well with the analytical results down to the fields of approximately 1,200 Vcm, and then start to deviate from the linear dependence and decrease faster for larger distances from the point $y = 0$. The calculation was performed for a temperature of 300 K and $Hg_{1-x}Cd_xTe$ with a composition of $x = 0.165$, which is optimum for that temperature.

Carrier concentration within the exclusion region is obtained by integrating the expression for dn/dy (3.83) and by utilizing the condition $\mu_p << \mu_n$, $n >> p$

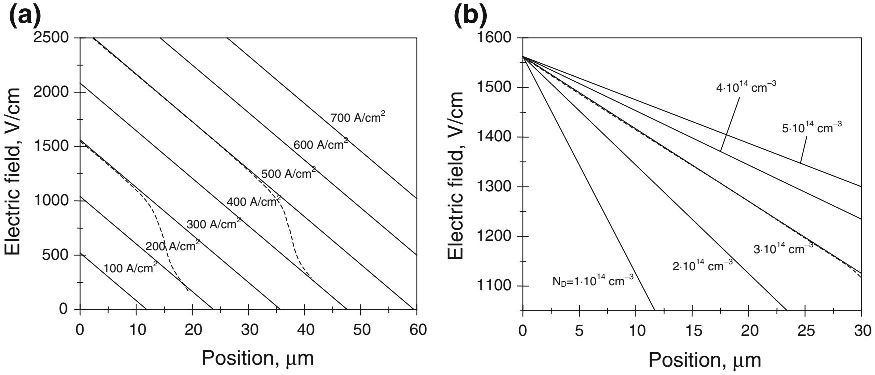


Fig. 3.7 Spatial distribution of electric field in an $n^+\nu$ exclusion detector. **a** Dependence on injected current density for a donor concentration of 3×10^{14} cm^{-3}; **b** dependence on donor concentration in excluded region for a detector current density of 300 A/cm^2. T = 300 K, material $Hg_{1-x}Cd_xTe$, composition x = 0.165. Dashed lines show electric field calculated numerically, solid lines are analytical values

$$n = n(0)\left(1 - \frac{g\mu_n p n(0) y}{\mu_p J |N_D - N_A|}\right)^{-1}, \tag{3.87}$$

where $n(0)$ is electron concentration for $x = 0$. This expression for majority carrier concentration is also applicable to conventional photoconductors with sweepout effect [50].

The further text presents a procedure to determine the total length of the region with significant exclusion processes. The above analysis is valid until a point is reached where the diffusion current is no longer negligible. This means that for the point of the exclusion zone end the diffusion term should be equal to a small percentage of the total current density. Let us denote that percent as r_w

$$r_w J = qD_n \frac{dn}{dy}. \tag{3.88}$$

By comparing the approximation with the results of the full numerical model White adopted for r_w a value of (1.2 ± 0.2) % [50]. In the point where (3.88) is valid for that value the numerically calculated electric field drops to 50 % of the value calculated according to (3.86).

If we replace the previously determined relation (3.83) for dn/dy into (3.88) and utilize the fact that the majority concentration and doping level are approximately equal, we obtain for the threshold value of the current at the edge of the exclusion region

$$J_{\text{thresh}} = qn(0)\sqrt{\frac{gD_n\mu_n}{r_w\mu_p|N_D - N_A|}}. \tag{3.89}$$

This means that there is a threshold current below which there is no exclusion. This is actually the minimum current necessary to decrease carrier concentration sufficiently for drift motion to prevail.

By replacing (3.89) into the simplified continuity equation (3.81) we obtain the critical electric field below which exclusion does not exist.

$$E_{\text{crit}} = \sqrt{\frac{gD_n}{r_w|N_D - N_A|\mu_n\mu_p}}. \tag{3.90}$$

The approximate expression for the total exclusion length is obtained by integrating the expression for the electric field derivative (3.85) from E to E_{crit}, where the length changes from 0 to L_{excl}:

$$L_{\text{excl}} = \frac{\mu_p J|N_D - N_A|}{\mu_n qgn(0)} - \sqrt{\frac{D_n|N_D - N_A|\mu_p}{gr_w\mu_n}}. \tag{3.91}$$

Figure 3.8 shows the dependence of the exclusion length on current density for various dopant concentrations.

In the case of a $p^+\pi$ type exclusion device, the expressions for the threshold current (3.89) and the exclusion zone length (3.91) are still valid [50]. The factor r_w is now approximately 0.08. An important difference in comparison to a n-type device is that now the parameters of the highly doped region influence the behavior of the device. The exclusion length also depends on the difference between the electric field in the point of maximum E_0 and the critical electric field.

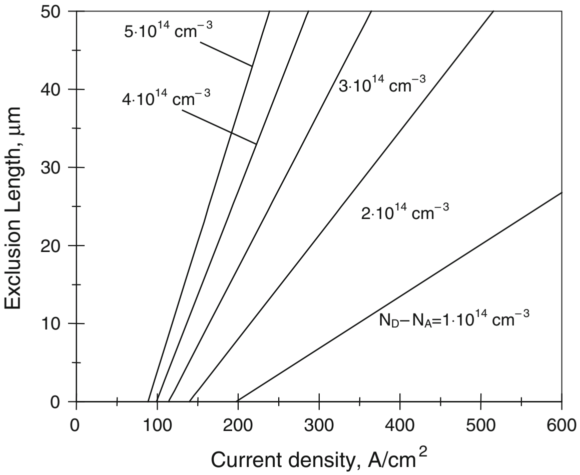


Fig. 3.8 Dependence of the exclusion length on injected current density in an $n^+\nu$ exclusion photoconductor for various concentrations of donors in the excluded region. T = 300 K, material is HgCdTe with a composition of 0.165

According to the approach from [50] we derive the expression for electric field in the point of majority concentration minimum for a p$^+\pi$ type exclusion device.

$$E_0 = \frac{J\mu_n\left(p_{p+} - n_{p+}\right)x_{p+} - qD_n n_{p+} p_{p+}\left(\mu_n + \mu_p\right)}{qx_{p+}\mu_n N_\Delta\left(n_{p+}\mu_n + p_{p+}\mu_p\right)}, \tag{3.92}$$

where the index p$^+$ means that the given variable is valid for the p$^+$ zone (y_{p+} is the width of the highly doped region, and n_{p+} and p_{p+} are the carrier concentrations within it).

Further we consider the sensitivity of an exclusion photoconductor. To this purpose we return again to the n$^+\nu$ device. By integrating the electric field according to (3.86) along the y-coordinate while adding the optical generation term (1.77) to the generation-recombination term we obtain the detector voltage. By further differentiating this expression for optical generation we arrive to an approximate expression for the sensitivity of exclusion detectors. It is necessary to take care that the exclusion length also changes with optical excitation. If the point y_1 where voltage is measured is within the exclusion region, the sensitivity of an n$^+\nu$ detector is

$$R_v = \frac{y_1\left(\mu_n + \mu_p\right)}{2hvwd\mu_n\mu_p|N_D - N_A}, \tag{3.93}$$

while for the same structure, if the measurement is performed across the whole device

$$R_v = \frac{\mu_n + \mu_p}{2hvwd\mu_n\mu_p|N_D - N_A|}\left(L_{\text{ekskl}} + \sqrt{\frac{D_n\mu_p|N_D - N_A|}{r_w g\mu_n}}\right). \tag{3.94}$$

White presented a generalization of the above expression [50] for the case of a gradient isojunction, where the concentration gradient is represented as $n = n_{n+}\exp(-x/a)$ (the parameter a describes the gradient nature of the concentration profile starting from the interface between the n$^+$ and the ν region). In this case the term $an_{n+}/2\tau_A^i g$ is added into the bracketed term of (3.94). This is an intuitively expected result, because it shows that the influence of the n$^+$ region is less felt in the detector performance if the isotype junction is more abrupt.

3.5.4 Carrier Distribution in Exclusion Detector

The further text gives a more detailed analysis of the properties of exclusion-based isojunction devices. The results were obtained by numerical simulation of devices. The full model for both carrier types was used.

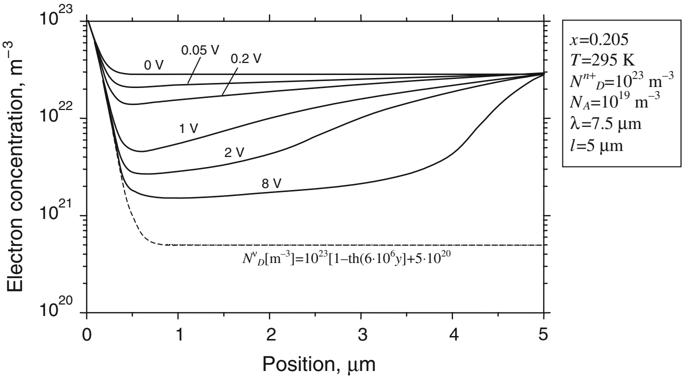


Fig. 3.9 Electron concentration profile along an $n^+\nu$ mercury cadmium telluride exclusion photoconductor with gradient isojunction for various values of reverse bias voltage (shown in figure) at 295 K. Dashed line shows donor concentration

Figure 3.9 shows electron concentration profiles, and Fig. 3.10 hole distribution for various reverse bias voltages for the case of an $n^+\nu$ exclusion photoconductor with gradient junction. The length of the whole structure was 5 μm, material HgCdTe with composition $x = 0.205$ at a temperature of 295 K. This composition is optimal for an IR wavelength of 7.5 μm.

The donor concentration in the n^+ region was $N_D = 10^{23}\ \text{m}^{-3}$, and the acceptor concentration $N_A = 10^{19}\ \text{m}^{-3}$. In the pseudo-extrinsic region the donor concentration was $N_D = 5 \times 10^{20}\ \text{m}^{-3}$. Starting from the isojunction the donor concentration changed along the device through the ν region as $N_D^\nu[m^{-3}] = 10^{23}\left[1-\tan\text{h}\left(6 \times 10^6 y\right)\right] + 5 \times 10^{20}$. The donor profile is shown as a dashed line in Figs. 3.9 and 3.10.

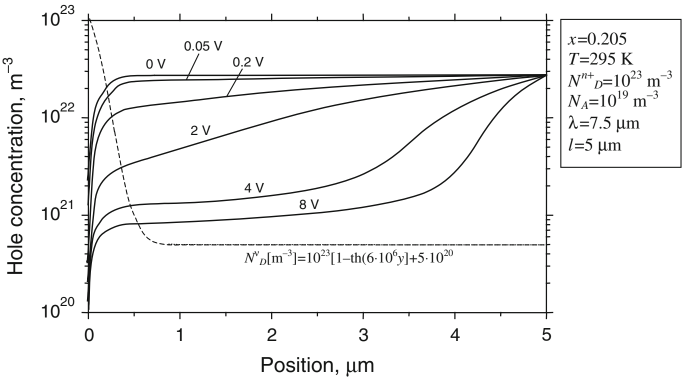


Fig. 3.10 Hole concentration profile along an $n^+\nu$ mercury cadmium telluride exclusion photoconductor with gradient isojunction for various values of reverse bias voltage (shown in figure) at 295 K. Dashed line shows donor concentration

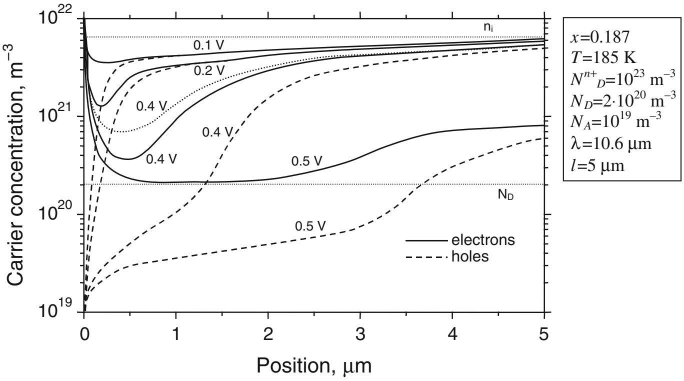


Fig. 3.11 Concentration profile for electrons (solid) and holes (dashed) along a mercury cadmium telluride exclusion photoconductor of n$^+$ν type with an abrupt junction at 185 K for various reverse bias voltages. Dotted line presents electron concentration at 0.4 Vm but with a donor gradient of the form of $N_D^{\nu}[\mathrm{m}^{-3}] = 10^{23}\left[1-\tan\mathrm{h}(6\times10^{6}y)\right]+5\times10^{20}$

It can be seen in Figs. 3.9 and 3.10 that no full exclusion occurs under the applied conditions, but only partial instead. Even for the largest values of bias the carrier concentrations are not constant, majority carrier concentrations are not brought near the impurity level and minority concentrations are not below it. Further simulations show that at room temperature exclusion does improve photoconductor characteristics, but cannot lead to qualitative changes and a level of Auger suppression sufficient to furnish a BLIP device.

Figure 3.11 shows the properties of an n$^+$ν with characteristics similar to those in Figs. 3.9 and 3.10, but at 185 K and with an abrupt junction. The device is again 5 μm long and the impurity profiles remain the same. The composition x = 0.187 is optimized for maximum sensitivity at 10.6 μm. It can be seen from Fig. 3.11 that the carrier exclusion is much larger and the necessary values of bias voltage much lower than in the previous case, although the cutoff wavelength is larger. Already at 0.5 V real exclusion occurs and the excluded region length is about 2.5 μm.

The main reason of the improvement of characteristics is a decreased operating temperature, but there is also a contribution from the shape of the transition region between the low-doped and the highly doped region, because “milder” transitions decrease the effects of exclusion. This can be seen if a comparison is made with the electron concentration curve obtained for the same dopant profile as in the case in Fig. 3.10. It is intuitively clear that the same effect will be caused by residual Auger processes (primarily Auger generation), and it is also visible from the analytical expression (3.87).

The electron and hole concentrations shown in Fig. 3.11 increase with coordinate with the same slope, but since logarithmic scale is used it appears that the slope of the electron curve is lower.

According to the results of the numerical calculations shown in Figs. 3.9, 3.10, 3.11, as well as according to the results of the approximate model, the exclusion length increases with electric field. For the lengths larger than this length electron and hole concentrations increase very rapidly (a faster-than-exponential growth) when the field collapse occurs, and approach the intrinsic concentration for the given operating temperature. The concentrations of majority and minority carriers are very near to each other in the partially excluded part. When the maximum exclusion is reached, the majority concentration remains very close to the extrinsic level.

3.5.5 g-r Rates in Exclusion Photodetector

Figures 3.12 and 3.13 show spatial distributions of Auger generation and recombination rates along an $n^+\nu$ mercury cadmium telluride exclusion device at a near-room temperature. The cadmium molar fraction was 0.205. The detector dimensions and doping profile were identical to those of the gradient structure described in Sect. 3.5.5.

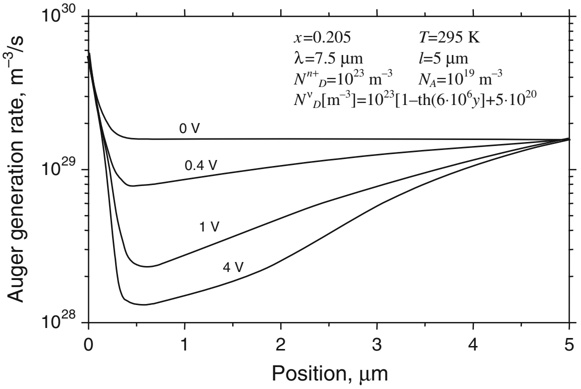


Fig. 3.12 Spatial distribution of Auger generation rate along the length of the exclusion region of an $Hg_{1-x}Cd_xTe$ $n^+\nu$ detector with a gradient junction at 295 K for various bias voltages

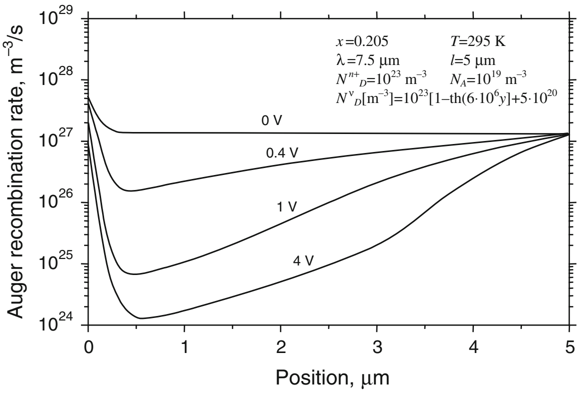


Fig. 3.13 Spatial distribution of Auger recombination rate along the length of the exclusion region of an $Hg_{1-x}Cd_xTe$ $n^+\nu$ detector with a gradient junction at 295 K for various bias voltages

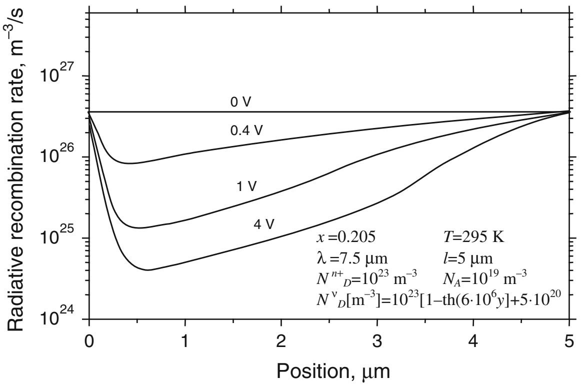


Fig. 3.14 Spatial distribution of radiative recombination rate along the length of the exclusion region of a $Hg_{1-x}Cd_xTe$ $n^+\nu$ detector with a gradient junction at 295 K for various values of bias voltage

In the part immediately near the n^+ region the processes of Auger generation (impact ionization) are reduced under the influence of applied bias of several volts for almost an order of magnitude, which is a relatively modest decrease. Further toward the end of the exclusion region the generation rate increases rapidly to reach finally in the bulk the level equal to that in unbiased structure. Auger recombination is suppressed to much lower levels and at a maximum exclusion its value is decreased about four orders of magnitude. The spatial profile of recombination rate follows at that the same qualitative shape as the generation curve.

Figure 3.14 shows radiative generation in the same structure. At the very beginning, at low values of reverse bias, its level is relatively low. The effects of carrier exclusion decrease it only about two orders of magnitude, so that Auger recombination falls below its level too.

Shockley-Read recombination rate is shown in Fig. 3.15. Its decrease is very modest, i.e., exclusion cannot be utilized for an efficient decrease of Shockley-Read processes.

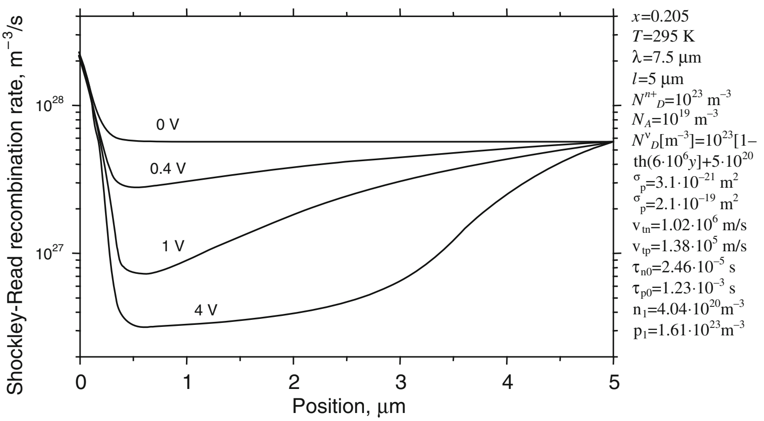


Fig. 3.15 Spatial distribution of Shockley-Read recombination rate along the length of the exclusion region of a $Hg_{1-x}Cd_xTe$ $n^+\nu$ detector with a gradient junction at 295 K for various values of bias voltage

According to the presented analysis, in an exclusion detector under sufficiently high bias the main limiting factor among the fundamental g-r processes remains the impact ionization.

3.5.6 *Current-Voltage Characteristics of Exclusion Photoconductor*

Current-voltage characteristics of mercury cadmium telluride exclusion devices are shown in Figs. 3.16, 3.17, 3.18. All of the presented curves are obtained for reverse bias of the isotype junction, but only the absolute values of current and voltage are shown in the diagrams. This is a convention used throughout this work.

Figure 3.16 shows the dependence of the current-voltage characteristics of an $Hg_{1-x}Cd_xTe$ $n^+\nu$ device designed to furnish its sensitivity maximum at 5 μm ($x = 0.27$) at an operating temperature of 270 K. The curves were calculated for various distances between the $n^+\nu$ junction and the point of measurement. It is readily seen that the UI characteristics are worse (necessary current density larger and dynamic resistance lower) for larger lengths.

Current-voltage characteristics reacts in a similar manner to an increase of the ν region doping (Fig. 3.17). All of the parameters in this diagram are identical to the previous one.

For a comparison, Fig. 3.18 shows the current-voltage characteristics of an $n^+\nu$ exclusion photoconductor at a temperature of 190 K. Several differences are readily noticeable: the exclusion threshold "knee" is much more clearly defined than in the previous case. The dynamic resistance is larger. Both of these results are a consequence of lower operating temperature. A negative property of the presented device is that the threshold current density dramatically increases with cutoff wavelength, i.e., the consequence is similar to the case of the dopant concentration increase.

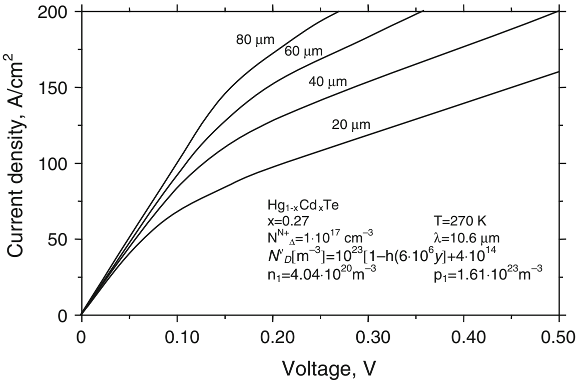


Fig. 3.16 Reverse current-voltage characteristics of an $n^+\nu$ exclusion device for various inter-electrode spacing at a temperature of 270 K

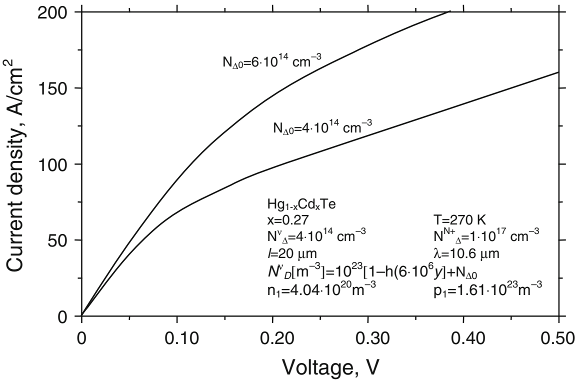
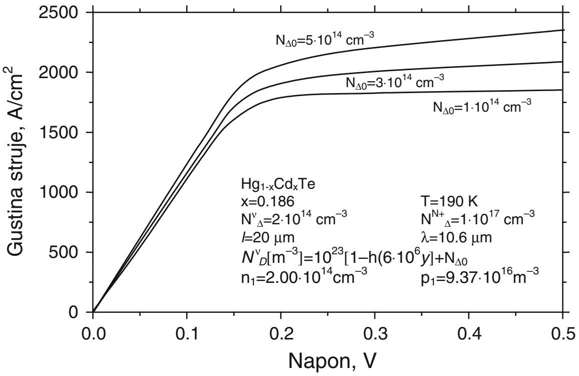


Fig. 3.17 Reverse current-voltage characteristics of mercury cadmium telluride $n^{+}\nu$ exclusion device for different dopant concentrations at a temperature of 270 K. Device parameters are shown in figure

Fig. 3.18 Reverse current-voltage characteristics of a mercury cadmium telluride $n^{+}\nu$ exclusion device for various dopant concentrations at a temperature of 190 K



None of the three presented current-voltage diagrams shows the effects of negative dynamic resistance. The slope of the characteristics in the beginning is defined by the Ohmic resistivity of material, i.e., by the ratio between the applied electric field and the term $(qn\mu_n + qp\mu_p)$. Further, when the ν region begins to be excluded, it crosses into a sub-Ohmic region (resistivity larger than according to the Ohm's law.

A conclusion is that the current-voltage characteristics progressively impairs (i.e., the exclusion threshold currents increase and dynamic resistance decreases) with an increase of the cutoff wavelength (i.e., with a decrease of the semiconductor bandgap), with an increase of the doping concentration in the ν region and with an increase of the exclusion length. It can be seen that the required nominal currents are very high, which is one of the most important drawbacks of exclusion devices. Negative differential resistance does not appear in any of the considered cases, not even for the shortest cutoff wavelengths and the lowest operating temperatures.

3.5.7 Calculated Noise of Exclusion Photoconductors

Noise current in exclusion detector was calculated utilizing the approach from Sects. 3.4 and 3.5. The structure was divided into infinitely thin slices along the y-axis with a thickness of d*y*, and then Eq. (1.94) was utilized to calculate the g-r component of noise. Photoelectric gain was determined according to (1.93). The g-r terms were determined according to (1.85) and (1.86), the optical generation term according to the simplified expression (1.80) and the total noise according to (1.95). The dynamic resistance was determined from the static characteristics for the case of a detector without illumination. Flicker (1/f) noise was neglected.

The presented results are valid for a mercury cadmium telluride $n^+\nu$ detector optimized for 10.6 μm. Figure 3.19 shows the calculated dependence of theoretical g-r noise current on detector bias current. Calculation parameters are shown in the figure. For low bias currents noise practically does not increase with the current, but this is the region where the exclusion effects are practically useless. A strong increase appears for higher bias values, while at still higher bias currents theoretical noise reaches saturation.

3.5.8 Sensitivity of Exclusion Photoconductor

Sensitivity of exclusion detectors, Fig. 3.20, was determined by calculating the dark current of detector according to the van Roosbroeck-Shockley model and including generation-recombination terms according to (1.85) and (1.86). The optical term (1.81) was subsequently added and the calculation was repeated. Signal current was

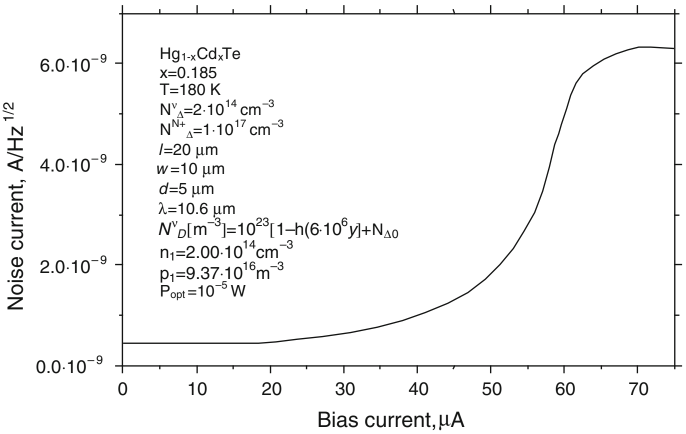


Fig. 3.19 Theoretical noise current of an exclusion photoconductor versus total bias current. T = 180 K, material HgCdTe, x = 0.185

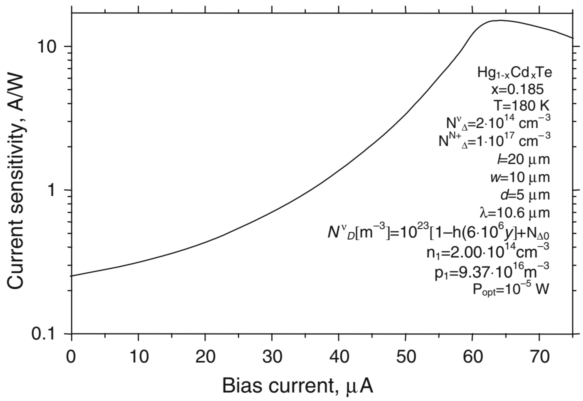


Fig. 3.20 Current sensitivity of exclusion detector versus bias current. $T = 180$ K, HgCdTe, $x = 0.185$

determined as the difference between the dark current and the current of illuminated detector. The optical power of signal was 20 μW in all of the cases.

The $n^+\nu$ $Hg_{1-x}Cd_xTe$ detector structure used to calculate current sensitivity was the same as in Fig. 3.19. It can be seen that the increase of sensitivity follows the same pattern as noise current. This is logical, since in situations when the g-r component prevails the same mechanisms are responsible both for creation of signal current and for its fluctuations.

It can be seen that at the highest values of bias a noticeable drop of sensitivity appears. This means that for each set of parameters of an exclusion device there is a value of bias furnishing maximum sensitivity.

3.5.9 Detectivity of Exclusion Photoconductor

The specific detectivity of an exclusion detector is shown in Fig. 3.21. It can bee seen that even the theoretical value of the obtained increase due to exclusion effect is modest (less than 2.5 times). The curve of the specific detectivity versus bias follows a dependence qualitatively very similar to that of current sensitivity.

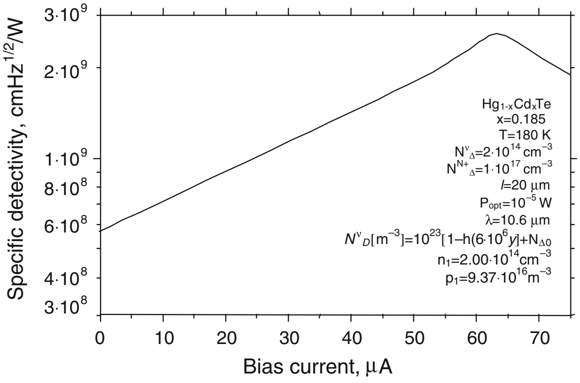


Fig. 3.21 Specific detectivity versus bias current for exclusion photoconductor, $T = 180$ K, material HgCdTe, $x = 0.185$

An interesting detail is that in logarithmic proportions the middle, rising part of the dependence is practically linear, meaning that there is a certain range of bias currents where specific detectivity increases exponentially. In real situations this is significantly impaired by the existence of 1/f noise.

For the largest values of bias a sharp drop of detectivity is seen, i.e., there is an optimum value for this dependence too.

A comparison between the calculated values of specific detectivity of an exclusion detector and the BLIP values for a field of view of 180° at the same cutoff wavelength (equal to performance of a photoconductive device at 77 K) shows that all of the calculated values are significantly below the BLIP limit. The detectivity increase in comparison to a photoconductor without nonequilibrium Auger suppression (conventional sub-BLIP photoconductor) is less three times.

3.5.10 Response Time of an Exclusion Photoconductor

In this section, an analytical consideration of transitional response of an exclusion photodetector is given. A starting assumption is that the response time is mostly defined by the drift transit time of photogenerated carriers across the active region. The influence of the detector circuitry is not taken into account.

The time of transit between the contact pads of a photoconductive device is determined as the ratio between the inter-electrode distance L and the product of electric field and the mobility of the considered carrier type, $T_n = L/E\mu$. If we divide the detector length into infinitesimally small elements dy, the differential transit time across each of these elements will be $\mathrm{d}t = \mathrm{d}y/E(y)\mu$. Thus the total transit time across the detector will be

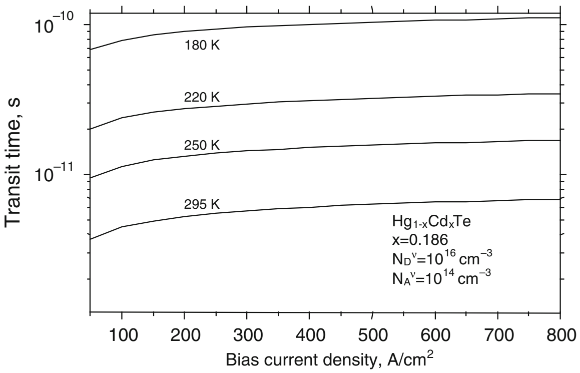
$$T_n = \int_0^{L_{\mathrm{excl}}} \frac{dy}{\mu E(y)}, \tag{3.95}$$

where L_{excl} is the length of the excluded region of the device under consideration, given by (3.92). The electric field $E(y)$ is given by (3.86) and its form is $a + by$. The integration of the expression on T_n gives a logarithmic dependence, and the replacement of the concrete values of the coefficients a, b and the exclusion length L_{excl} furnishes

$$T_n = \frac{N_D - N_A}{g} \ln\left[1 + L_{\mathrm{ekskl}} \sqrt{\frac{r_w \mu_n g}{D_n \mu_p (N_D - N_A)}}\right]. \tag{3.96}$$

The transit time between some other two points, determined by the position of the signal readout contacts that may in a general case differ from the bias contacts, is obtained in an almost identical manner, by integrating the expression on T_n within different limits.

Fig. 3.22 Transit time in an exclusion photoconductor versus operating temperature for different dopant concentration. Material $Hg_{1-x}Cd_xTe$, $x = 0.186$



Expression (3.96) can be represented as [50]

$$T_n = \frac{N_D - N_A}{g} \ln \frac{J}{J_{th}}. \tag{3.97}$$

Figure 3.22 shows a dependence of the transit time of a mercury cadmium telluride exclusion photoconductor on bias current density for different temperatures. The cadmium molar fraction was kept constant in all cases, i.e., the compositions are not constant for the given temperatures.

It can be seen that shorter transit times are obtained for higher temperatures, which is a desirable property, since the basic application of nonequilibrium devices is to elevate the operating temperature of photodetectors. On the other hand, it appears what may seem paradoxical that a bias increase influences detrimentally the transit time. Actually the electric field increases carrier velocity, but at the same time it increases the length of the exclusion zone, so that a carrier takes longer time to cross the whole zone. A transit across a part of the excluded zone with a constant length would, naturally, show an increasingly shorter transit time with a bias increase.

3.5.11 Experimental Results on Excluded Nonequilibrium Photodetectors

Exclusion of minority carriers in narrow-bandgap semiconductors was demonstrated in a temperature range of 180–295 K, exclusively in the middle wavelength infrared range, i.e., in the range (3–5) μm. Most of the reported experiments were performed on ν-type semiconductors [243].

Ashley et al. [333] fabricated a number of exclusion structures in mercury cadmium telluride with compositions ranging from 0.23 to 0.30 and for various cutoff wavelengths. A typical structure was 2 mm long, 0.1 mm wide, and 4–8 μm thick. Highly doped n^+ zone was usually formed by ion implantation, with a typical doping level of 5×10^{17} cm^{-3}. The active area was passivated by zinc sulfide that served at the same time as a single-layer antireflection coating. An array of contacts was formed along the detector serving to measure UI characteristics for different positions along the exclusion zone. Carrier concentration in material was determined by Hall measurements using a magnetic induction of about 14 mT.

Large nonequilibria were reached at temperatures above 200 K. The effects of carrier heating were felt at near-room temperatures. Large sensitivity improvements were measured in the reverse bias mode in comparison to an equivalent directly biased structure, of the order of 25 times. Spectral noise was also measured in a range of 1–100 kHz. It was found that the level of flicker noise was excessively high. It was also noted that this type of noise increased strongly with an increase of bias current. This was the reason why the total measured detectivity improvement was only 2 to 3 times.

Djurić et al. [358] fabricated InSb-based exclusion structures with light signal arriving through the n^+ zone, rather than laterally as in the quoted structures of Ashley et al. [333]. The highly doped region was formed by implantation of tellurium, with a thickness of (5–10) μm. The level of the ν-substrate doping was 10^{14} cm^{-3}. An exclusion threshold current of 270 $\mathrm{A/cm}^2$ was measured and the maximum temperature with measurable exclusion effects was 220 K.

The largest problem in all experimental exclusion structures was an excessive level of 1/f noise. Some experiments were performed with a goal to reduce this noise by using the procedure of correlated double sampling [333]. To this purpose the authors fabricated two lateral contact pairs on the detector to read the voltage signal. The part of the detector leading to one of these pairs is screened by an opaque mask, so that only noise is read in that contact pair. The part of the detector corresponding to the second pair is illuminated by signal. The phase difference between these two pairs is negligible if the fluctuations of carrier concentration occur in a period longer than the transit time between the first and the second contact pair. On the other hand, the sampling time will very probably be sufficiently long, since the method is utilized to eliminate low-frequency noise.

If the readouts from both of the contact pairs are lead to separate preamplifiers, brought into counterphase and lead to an addition circuit, pink noise will be removed, since it is correlated on both channels, but the signal will remain, since it exists on only one of the contacts.

This method was applied to exclusion devices and an attenuation of 1/f noise of about six times was obtained [333]. One of its drawbacks is that it is necessary to perform additional signal adjustment in one of the branches (for instance using variable attenuator) to compensate possible differences in noise intensities between the branches.

3.5.12 Advantages and Shortcomings of Exclusion-Based Nonequilibrium Photodetectors

Technologically and physically exclusion devices are surely the simplest Auger-suppressed devices. Their operation for the most part corresponds to a conventional photoconductor in the “sweepout” mode. The only required additional measure to ensure their operation is protection against overheating, and this is a problem with all nonequilibrium devices generally.

The main hindrance to a more widespread use of exclusion devices are large bias current densities required for operational mode. Depending on the designed cutoff wavelength they range from a few hundreds to several thousands A/cm^2. Any increase of operating wavelength dramatically increases dark current. Among impeding factors are also the use of semiconductor single crystals with a larger number of defects, as well as temperature increase.

At the wavelengths required for night vision the required values of bias are so large that they enter the hot electron region, since the limiting values of the allowed electric fields are of the order of several hundreds of V/cm. A comparison of critical electric fields required for the start of exclusion with the maximum allowed values of bias where the effects of hot carriers are not prevailing shows that near-BLIP exclusion devices are possible at room temperature in the wavelength range of (3–5) μm, while in the range of (8–14) μm the operating temperature would have to be below 200 K—all of this under the assumption that the carrier concentration levels are about 10^{14} cm^{-3}.

On the other hand, the large current densities pose the problem of Joule heating and excess heat removal from the active area. One of the proposed solutions is a fabrication of the smallest possible detector structures. For instance, for a current density of about 2,000 A/cm^2 with a detector about 10 μm thick and 50 μm wide its total dissipation would be about 5 mW.

Another large problem with these detectors is its excessive flicker noise, additionally increased with bias. This is a very important shortcoming, because to avoid noise introduced by Auger processes another kind of noise is introduced, proportional to bias current. Indeed, experimentally measured noise sensitivity improvements in the first experiments with exclusion devices were at least an order of magnitude [333], but the obtained detectivity improvements were very modest in comparison, about a factor two. An additional problem was that the measurement frequency had to be high enough to avoid 1/f component.

Because of the above reasons, the exclusion devices are today completely abandoned as a stand-alone solution for the suppression of Auger g-r processes. And yet, in this consideration they received a lot of attention. The reason for this is purely practical. The principles and approaches connected with this type of devices are directly or indirectly used for all of the other known types of nonequilibrium detectors. Besides this, the exclusion isojunction is one of the two basic elements of the structure of all photodiodes with extraction-based suppression of Auger processes, as it will be presented in the next Sect. 3.6.

3.6 Extraction Photodiodes

The most widely used effect for Auger process suppression in infrared MWIR and LWIR photonic detectors is minority carrier extraction. This is a phenomenon appearing on reversely biased p-n junction, i.e., it is connected with diode-type photodetectors. These detectors are denoted as “Auger-suppressed photodiodes”, “extraction diodes”, and “detectors with carrier extraction”. Strictly speaking, these are extraction-exclusion devices, i.e., they simultaneously utilize the effects of minority carrier exclusion and extraction.

Minority carrier extraction can be described as follows (e.g., [360]). Let us consider a schematic presentation of carrier distribution in a p-n junction (Fig. 3.23).

It is well known that an unbiased p-n junction creates a depleted region by balancing diffusion and drift by a built-in electric field. By applying reverse bias the potential barrier is increased by the amount of the applied bias voltage, while the number of majority carrier with an energy large enough to cross to the other side of the p-n junction decreases proportionally to exp($-qU/k_bT$), i.e., exponentially (U is applied bias). This means that at sufficiently high reverse voltages ($qU \gg k_bT$) the majority current becomes negligible. At the same time, the opposite flow (leakage of minority carriers across the p-n junction where they become majority) is not hindered in any way.

If we assume that the applied electric fields are low enough, according to the continuity equation the electron spatial distribution will be

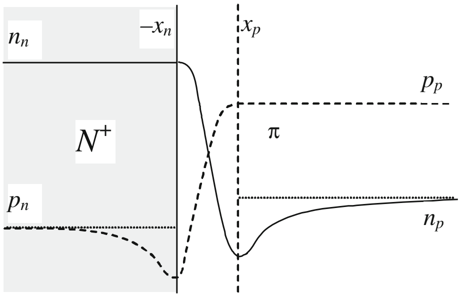
$$n = n_p\left[1 - e^{\frac{x_p - x}{L_n}} + e^{\frac{U}{k_b T}} e^{\frac{x_p - x}{L_n}}\right]. \tag{3.98}$$

Here L_n denotes electron diffusion length, much shorter than the width of the quasineutral region in p-type semiconductor. If the p-n junction is reversely biased so that $qU/k_bT \gg 1$, the electron concentration becomes

$$n = n_p\left[1 - e^{\frac{x_p - x}{L_n}}\right]. \tag{3.99}$$

This means that in a region with a width of approximately L_n near the boundary of the depleted layer the concentration of minority carriers (electrons in the case under consideration) will decrease, where the decrease will depend on the value of the diffusion length of minority carriers, i.e., on the recombination parameters of the p-region.

Fig. 3.23 Spatial distribution of carriers in a p-n junction



More accurately, we may write

$$n = n_p\left[1 - e^{\frac{x_p - x}{L_n}} + e^{-\frac{|U|}{k_bT}}e^{\frac{x_p - x}{L_n}}\right]. \tag{3.100}$$

At a distance of $x \approx x_p$ the electron concentration becomes

$$n = n_p e^{-\frac{|U|}{k_bT}}. \tag{3.101}$$

The phenomenon of a sharp decrease of minority carrier concentration due to diffusion at a distance of L_n if the p-n junction is reverse biased is called extraction (reverse of injection).

If an exclusion isojunction layer is situated near the extraction junction at a distance larger than L_n, the two junctions will decrease the carrier concentration in the detector independently on each other. The situation dramatically changes if these two junctions are situated at a distance smaller than the minority carrier diffusion length. In that case their effects add up and they jointly decrease minority carrier concentration by several orders of magnitude, down to levels beyond reach of each of these mechanisms separately. The majority concentration at that tends to the impurity level.

The basic types of extraction structures are $p^+\pi n^+$, $p^+\nu n^+$, PνN and PπN, as well as their combinations $p^+\pi$N, $P^+\pi$N, etc. The middle (active) region (π or ν) is low doped, meaning that the diffusion length in it is large. Since the thickness of the active area must be smaller than the diffusion length, it means that it is more convenient to use a π than a ν active region. Since the diffusion length of minority carriers in a ν region is much smaller, it would be necessary to use very thin layers. Any thickness of the active area larger than the diffusion length of minority carriers only contributes to leakage currents and elevates the noise level. The contact regions must be thick enough and very highly doped (this is especially valid for p^+) so that their conductivity is low and the internal electric field low.

The performance of the diode structures utilizing the effects of minority carrier extraction is much better than that of the exclusion photoconductors. Their leakage currents are much lower, and Auger suppression much larger at significantly lower electric fields.

3.6.1 Historical Development of the Use of Minority Carrier Extraction in Semiconductor Detectors

The phenomenon of minority carrier extraction at a barrier junction as a process reverse to the injection has been known from the very beginnings of the research investigation of semiconductors, the injection itself being one of the basic mechanism of operation of semiconductor devices.

Banbury noted in 1953 a conductivity drop in PbS samples caused by extraction [354]. Arthur et al. proposed in 1955 the use of extraction to intentionally cause large changes of carrier concentration [354] and suggested a possibility to fabricate new semiconductor devices based on minority carrier extraction.

Ashley and Elliott were the first to propose nonequilibrium infrared devices with Auger process suppression using minority carrier extraction [338]. Their devices were $p^+\pi n^+$ and $p^+\nu n^+$ diodes fabricated in narrow-bandgap semiconductors and simultaneously utilizing the effect of exclusion and extraction. Due to the complexity of the proposed structures the first experimental models were only fabricated after several years, while their development is not finished even today.

The first extraction structure for Auger suppression was fabricated in InSb [336]. It was a three-layer homojunction structure with two contacts. An extraction diode in mercury cadmium telluride was fabricated as a three-terminal device utilizing the pseudoexclusion effect [361].

Four-layered structures of the $p^+P^+\pi n^+$ type (barrier type) fabricated in InAlSb were reported in 1990 [336], and five-layered (fabricated by metalorganic epitaxy) in 1996 [362]. Transistors utilizing the method of extraction and operating at elevated temperatures were described in [363]. They were arranged in a comb-like structure made of densely packed diodes that together functioned as an npn transistor. In 1991 four-layered n-channel MISFET structures were reported, as well as narrow-bandgap diodes intended for operation at room temperature [363]. In 1994 InAlSb emission structures were implemented with radiation wavelengths of 6 μm and 7 μm [364], and in 1996 HgCdTe emitters were reported for the wavelength range up to 10 μm [365]. Emelie et al. performed optimization of extraction photodiodes utilizing finite element modeling [366, 367].

3.6.2 *Various Implementations of Nonequilibrium Photodetectors Utilizing Extraction and Exclusion*

The basic and the oldest type of extraction-exclusion diode structure is shown in Fig. 3.24. It is a three-layered, two-contact structure and from its improvements, and development stemmed all other existing detector versions.

The structure was first fabricated in a "vertical" layout [363]. The middle layer (active region) may be fabricated in near-intrinsic p-type semiconductor (π) or possibly ν-type (in that case the exclusion contact is below, and the extraction contact above). The thickness of this layer must be smaller than the diffusion length of minority carriers (e.g., 1.5 μm, about one third of the absorption length) so that the exclusion and the extraction junction may function as a single unit. A typical thickness of the n^+ or of the p^+ region is about 10 μm.

Figure 3.25 shows the second generic type of extraction devices. This is a diode with three or more contacts [363]. The contact on the left side is purely Ohmic, while the right one is of $n^+\pi$ type. Here G denotes a "protective" $n^+\pi$ diode whose

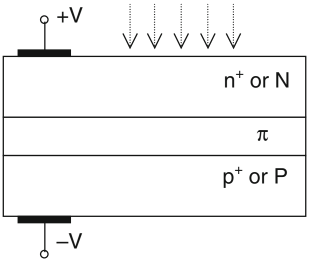
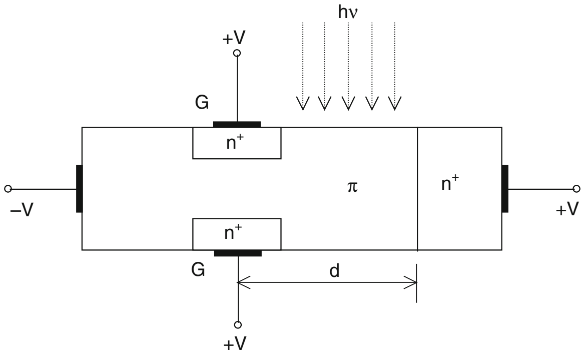


Fig. 3.24 Three-layer, two-contact $p^+\pi n^+$ or PπN extraction diode

Fig. 3.25 Extraction diode with auxiliary “pseudo-exclusion” contacts, i.e., utilizing “proximity extraction”



role is similar to the guard ring in conventional pin structures. When reversely biased, it simulates exclusion by drawing away the electron current injected by the Ohmic contact, which is the reason why the authors of the structure called its influence pseudo-exclusion. The effect itself was called “proximity extraction”.

The distance within the π region from the $n^+\pi$ junction to the narrowing (“constriction”) denoted in Fig. 3.25 as *d* must be much smaller than the diffusion length of electrons. The concentration of minority carriers within the region *d* is very low, and starting from the narrowing toward the Ohmic contact it increases until it reaches intrinsic concentration n_i for the given temperature.

In the experimental structures of the above type the achieved reductions of leakage current density reach the values tenfold smaller than in two-contact exclusion structures.

To avoid the need for an extremely high doping of the p^+ region required to suppress Auger 7 g-r processes (prevailing in p-type narrow-bandgap materials near-room temperature) four-layered structures with heterojunction contacts were introduced [363]. One such nonequilibrium photodiode is shown in Fig. 3.26. In this structure a highly doped strained layer made of wide-bandgap material P^+ is inserted between the p^+ and the π region. The thickness of this layer must be larger than the tunneling length. Its thickness is limited by the strain of the crystal lattice (for instance when using InAlSb on InSb).

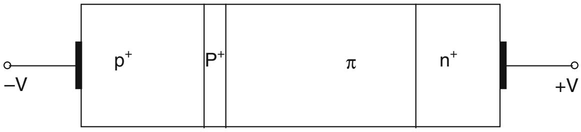


Fig. 3.26 Four-layered extraction-exclusion layer nonequilibrium IR diode incorporating a wide bandgap barrier layer

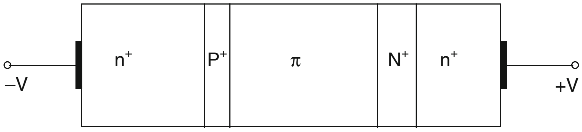


Fig. 3.27 Five-layered photodiode with Auger suppression incorporating two wide bandgap barrier layers

A typical thickness of the P^+ layer is between 10 and 20 nm. In the conduction band it creates a barrier with a height much larger than kT and thus decreases the electron diffusion length [363].

Figure 3.27 shows a further generalization of the above solution, a nonequilibrium photodiode with five layers.

Nowadays an important position belongs to the infrared detectors incorporating the mentioned wide-bandgap semiconductor barrier [368–370] (the BIRD detector —Barrier Infrared Detector). Maimon and Wicks proposed to use unipolar BIRDs —i.e., the built-in barrier layer blocks one carrier type, but allows free flow of the other type [371]. Such structures are for instance nBn. Ting et al. proposed the use of complementary barriers, one for electrons, another for holes, positioned at different depths [372]. Itsuno et al. analyzed NBvN and nBn detectors (where B stands for Barrier) [373, 374]. In their 2013 paper Martinyuk et al. quoted that besides Auger suppression the BIRD devices also suppress Shockley-Read-Hall g-r processes [375].

The whole concept of extraction photodiodes poses a practical problem, since the requirement exists that the active region thickness must be smaller than the diffusion length. Such a requirement poses design restrictions regarding the absorption path and the total detector active area. One of the approaches utilized to solve the above problem was not to use a single detector but rather a densely packed matrix of separate elements with lateral contacts on a joint substrate and with a joint exclusion junction (Fig. 3.28a) [376].

The matrix structure has a two-dimensional array of equidistant extraction contacts with a diameter smaller than the diffusion length. There is a positive contact on each of the elements, while the whole region is surrounded by a ring with negative contact. Breaking the extraction contact into a large number of elements enables an unlimited spreading of the region where the extraction effect appears. An additional advantage is that the signal from each of the n^+ elements can be independently registered, so that a single image pixel corresponds to it.

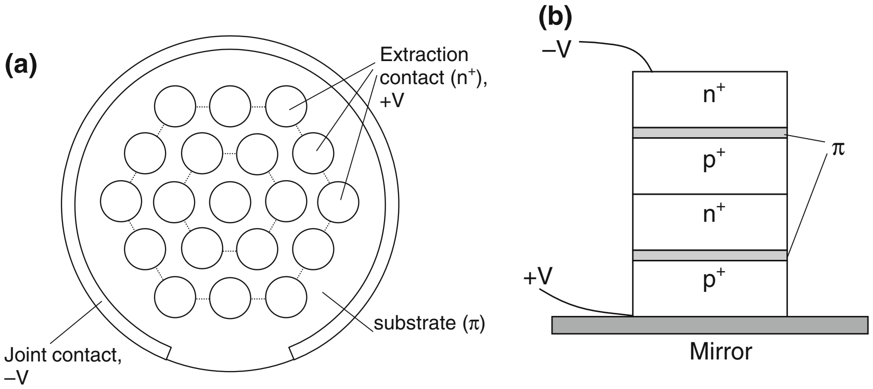


Fig. 3.28 Methods for the improvement of absorption path and total dimensions of extraction photodetectors. **a** Matrix of extraction diode elements of $n^+\pi$ type on a joint substrate; **b** multilayer "sandwich" structure on reflective substrate

A similar structure was used when designing npn transistors [377]. In this case cylindrical rings were formed instead of the n^+ field, so that each pixel becomes one diode. The middle diode is then the emitter, the first ring around it is the collector, the next one is the second collector, etc.

The second structure is a "sandwich" of several detectors, shown in Fig. 3.28b [376]. Such a structure enables an effective increase of the active layer thickness (fourfold in the case presented in the figure) without compromising carrier transport properties of the extraction-exclusion device.

3.6.3 Approximate Analytical Calculation of Narrow-Bandgap Extraction-Exclusion Photodiodes with Abrupt Junction

In the further text we consider a three-layer extraction-exclusion PπN structure intended for operation at an operating temperature of (200–300) K. The presented analysis can be also applied to the case of "high-low" junctions. In the active region of the considered structure electrons and holes are present in approximately equal concentrations. Thus, there will be a strong coupling between them resulting in marked ambipolar effects. Also, similar to all other types of nonequilibrium Auger-suppressed devices, deviations from equilibrium will be very marked (especially near the junctions) so that only the large signal theory will be applicable.

The one-dimensional continuity equation for the ambipolar transport of charge carriers will be posed as [378]

$$\frac{\mathrm{d}}{\mathrm{d}y} D_a \left(\frac{\mathrm{d}n}{\mathrm{d}y} \right) - \frac{\mathrm{d}}{\mathrm{d}y} (-\mu_a E) + g = 0. \tag{3.102}$$

where the coordinate system is positioned so that $x = 0$ in the point of the Nπ junction, while the isotype junction is in $y = -w$. We assume that w is much smaller than the diffusion length of minority carriers.

The ambipolar coefficient of drift mobility is equal to $\mu_a = \mu_n \mu_p (p - n)/(\mu_n + \mu_p)$. Since in the depleted region the condition $n - p = \text{const} \ll n, p$ is valid, the drift term can be neglected. The continuity equation then becomes

$$D_a \frac{d^2 n}{dy^2} + g = 0. \tag{3.103}$$

Boundary conditions are

$$\begin{aligned} n &= n(0) \\ \left.\frac{dn}{dy}\right|_{y=-w} &= 0, \end{aligned} \tag{3.104}$$

where $n(0)$ is some initial value of electron concentration in the point y = 0, while w is the width of the extraction region.

The total current of the device must be constant in all of the points y. In the position of the p-n junction ($y = 0$) it consists solely of the electron current ($J_p = 0$), and on the isotype junction ($y = -w$) it consists solely of the hole current ($J_n = 0$). Therefore

$$J = 2 D_n q \left.\frac{dn}{dy}\right|_{y=0}, \tag{3.105}$$

$$J = -2 D_p q \left.\frac{dn}{dy}\right|_{y=-w}, \tag{3.106}$$

where the condition was used that in the depleted region the equality $\mathrm{d}n/\mathrm{d}y = \mathrm{d}p/\mathrm{d}y$ must be valid.

Ashley and Elliott [333] neglected recombination terms in the depleted part and left only Shockley-read and radiative generation to propose the following expression for electron concentration in the π region

$$n(y) \approx (G_{SR} + G_{\mathrm{RAD}}) \tau_A^i \left(\cos\left(\frac{y + w}{L_a} \right) \cos\mathrm{ec}\left(\frac{w}{L_a} \right) - 1 \right) n(0), \tag{3.107}$$

where L_a is the ambipolar diffusion length denoted as $L_a = \left(D_n \tau_A^t \right)^{1/2}$.

The usual expression for reverse current density in a diffusion limited diode is (e.g., [58])

$$J = J_s \left\lfloor 1 - e^{-qU/k_bT} \right\rfloor . \tag{3.108}$$

In a photodiode with Auger suppression the saturation current density decreases with reverse bias, thus leading to the appearance of negative differential resistance. For the above described case we calculate this current according to

$$J_s = qD_n \frac{\partial n(y)}{\partial y} \bigg|_{y=0} \tag{3.109}$$

so that

$$J_s = q(G_{SR} + G_{\mathrm{RAD}})L_a \, tg \frac{w}{L_a} \tag{3.110}$$

Figure 3.29 shows the dependence of saturation current density on operating temperature for various compositions of the active region in mercury cadmium telluride extraction diode. The dependence was calculated according to (3.110), where (1.85) was used to determine the g-r term with neglected Auger and SR generation, and L_a was calculated as $L_a = (D_n \tau_{\mathrm{ai}})^{1/2}$. The intrinsic Auger lifetime was determined according to the Beattie-Landsberg expression (1.51). The width of the active region was $w = 2$ μm.

The electric field is estimated according to the condition that the total current density through the device must be constant. If we use $n = p - N_\Delta$, then by applying (3.105) and (3.106) and using the expression for the current density (3.110)

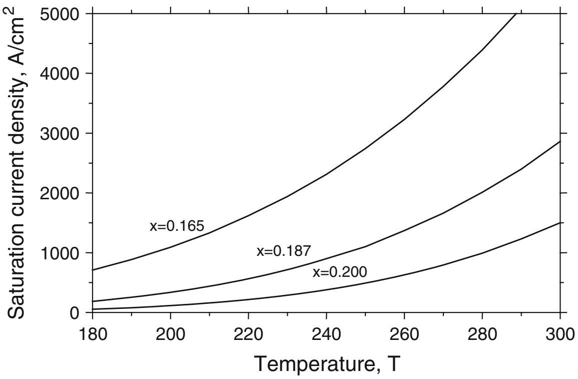


Fig. 3.29 Saturation current density versus operating temperature for various compositions of the active area in a mercury cadmium telluride extraction photodiode limited by radiative processes. The active region width $w = 2$ μm

$$E = \frac{J_s}{q\mu_p N_\Delta}. \tag{3.111}$$

$$E = (G_{SR} + G_{\mathrm{RAD}})\frac{w}{\mu_p N_\Delta}, \tag{3.112}$$

The approximate expression for the detectivity of an extraction-exclusion device will be [333]

$$D^* = \frac{\eta}{\hbar\omega}\sqrt{\frac{q}{2J_s}}. \tag{3.113}$$

and thus

$$D^* = \frac{\eta}{\hbar\omega}\sqrt{\frac{ctg(w/L_a)}{2L_a(G_{SR} + G_{\mathrm{RAD}})}}. \tag{3.114}$$

3.6.4 Carrier Distribution in Extraction-Exclusion Photodiode

The spatial distribution of carrier concentration in extraction structures was calculated using the general model of nonequilibrium photodetectors under the influence of purely electrical field (3.65)–(3.69) with the boundary conditions (3.70)–(3.74). The generation term was calculated according to (1.85), (1.86). One of the convenient methods of calculation is to use finite difference method (e.g. [379]).

Figures 3.30 and 3.31 show the concentration profiles in a three-layer HgCdTe extraction-exclusion $p^+\pi n^+$ photodiode at a temperature of 250 K for various values of bias voltage. The active region width was 2.7 μm, the p^+ layer was 0.65 μm wide, the n^+ layer 0.5 μm. Material composition throughout the whole diode was $x_{\mathrm{Cd}} = 0.187$. Donor concentration in the n^+ region was 10^{18} cm^{-3}, and the concentration of acceptors in the p^+ layer had the same value. The p-dopant concentration in the active region was 10^{14} cm^{-3}. Both the isotype and the p-n junction were assumed to be abrupt.

The obtained values for the concentrations of both carrier types in the most part of the active area are very close to each other. More significant differences can be noted only at the edges of the region. The electron concentration is slightly below the hole concentration. An increase of reverse bias voltage decreases the concentration of both carrier types below the equilibrium level in the π region (dashed line).

Calculations for $p^+\nu n^+$ structures give results almost identical to those for $p^+\pi n^+$ photodiodes with the same doping levels. The main difference is that here the level of electron concentration is slightly above the hole level.

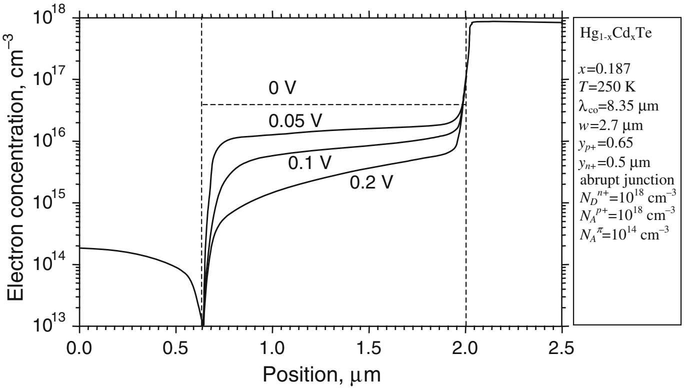


Fig. 3.30 Calculated spatial distribution of electron concentration in an extraction $p^+\pi n^+$ photodiode in mercury cadmium telluride. The values of the applied reverse bias are given near the corresponding curves. The horizontal dashed curve denotes the carrier concentration at zero bias (intrinsic concentration)

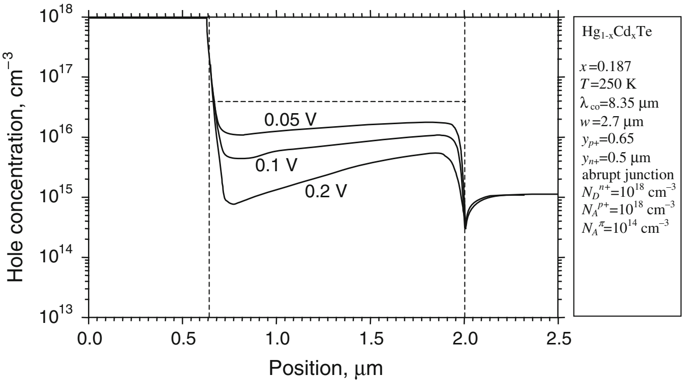


Fig. 3.31 Calculated spatial distribution of hole concentration in an extraction $p^+\pi n^+$ photodiode in mercury cadmium telluride at a temperature of 250 K, the active area width 2.7 μm. The horizontal dashed curve denotes the carrier concentration at zero bias (intrinsic concentration)

According to the analysis of the behavior of extraction devices published in [380, 381] an increase of the active region thickness in an extraction-exclusion detector, independently on the doping level, decreases the depletion of both types of carriers and impedes exclusion and extraction processes, i.e., in both cases there is a maximum thickness of the active region. This limiting value of thickness is approximately equal to the diffusion mean free path length $\left(2D_p\tau_A^{i}\right)^{1/2}$. In the case when this limit is too small, the device actually behaves as two independent photodetectors connected in series, one of them an extraction device and the other an exclusion detector. In the case when the active region is extrinsic at the operating temperature without bias the depletion of carriers becomes very sensitive to the value of the main type of dopants. In an n-type active region the depletion is very weak.

The second device type is the photodiode whose active region is of p-type when there is no bias. In such devices a reverse bias causes much larger depletion of carriers, and the necessary voltages are much lower. The diffusion length reach several tens of micrometers (because of the larger diffusion coefficients of electrons—minority carriers.) Electron concentration may become very low if the p-doping level in the beginning was such that already the initial electron concentration was low.

A conclusion is that in homojunction extraction photodiodes whose active area is always extrinsic it is much more convenient to use p-type doping, while the performance becomes approximately equal in pseudoextrinsic active regions (extrinsic only after bias is applied).

In the case of heterostructure junctions (the structures of NiP type, where “i” stands for a pseudoextrinsic ν or π region) the material parameters change a both junctions. Mobility and diffusion coefficient become coordinate-dependent, the same as the absorption coefficient and the refractive index. A convenient choice of the concrete values may lead to a significant improvement of carrier depletion and to a significant decrease of leakage currents as a consequence.

3.6.5 g-r Rates in Extraction-Exclusion Photodiodes

Figure 3.32 shows the dependence of Auger recombination rate on position within a $p^+\pi n^+$ type extraction photodiode (the corresponding carrier profiles are shown in Figs. 3.30 and 3.31), and Fig. 3.33 shows Auger generation in the same device. Similar to exclusion devices, Auger recombination suppression is much larger than Auger generation suppression. The largest suppression is in the part of the active region near the isotype junction, and contrary to the case of exclusion devices, Auger generation, and recombination rates suppression strongly increases already at low electric fields. With a value of bias voltage of 0.2 V Auger recombination suppression is over three orders of magnitude, and about two orders of magnitude for generation. Auger generation remains approximately constant outside the detector active area.

Figure 3.34 shows the distribution of radiative recombination rate across an extraction-exclusion device. Radiative generation is constant throughout the whole presented homojunction structure, since it is dependent only on material parameters (constant throughout the structure) and detector geometry, and its calculated value is 3.46×10^{27} m^{-3}/s.

In the case of double heterostructure photodiodes the level of radiative generation is position dependent and reaches the highest value within the active region.

While calculating the radiative processes rate the influence of reabsorption effects was taken into account. A reflection coefficient of 0.1 at the front surface and 1 at the back surface were assumed.

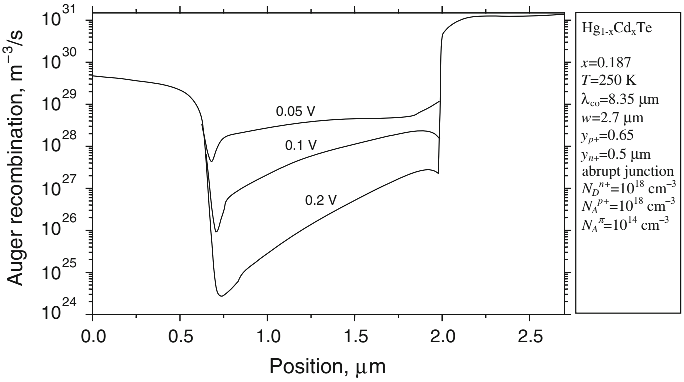


Fig. 3.32 Spatial dependence of Auger recombination rate in a reverse biased extraction-exclusion HgCdTe photodiode at 250 K for different values of bias voltage

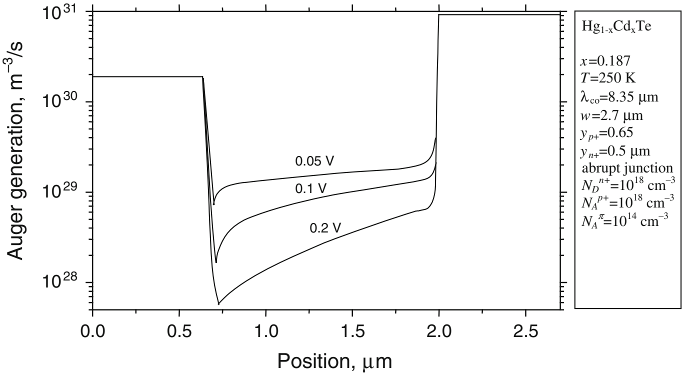


Fig. 3.33 Spatial dependence of Auger generation rate in a reverse biased extraction-exclusion HgCdTe photodiode at 250 K for different values of bias voltage

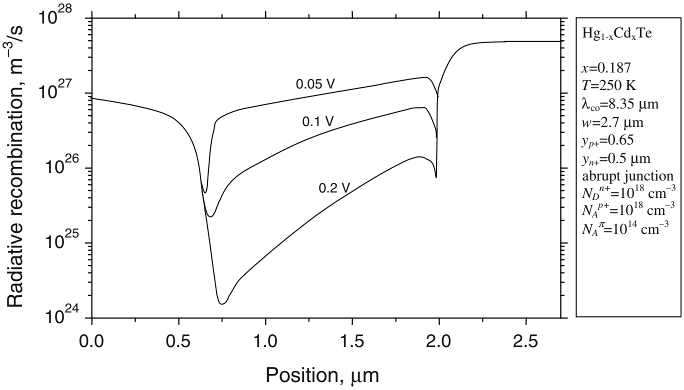


Fig. 3.34 Spatial dependence of radiative recombination rate in a reverse biased extraction-exclusion HgCdTe photodiode at 250 K for different values of bias voltage

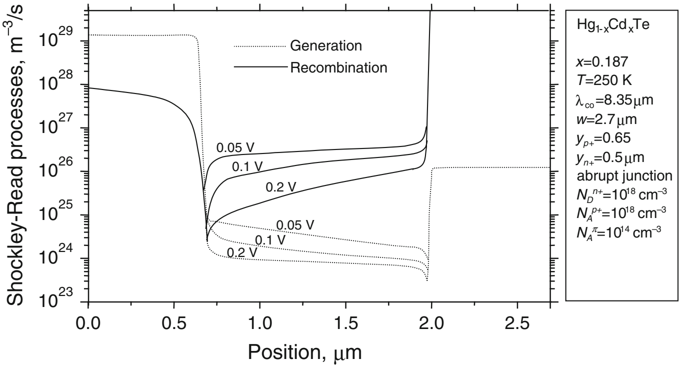


Fig. 3.35 Spatial dependence of Shockley-Read g-r rates in a reverse biased extraction-exclusion HgCdTe photodiode at 250 K for different values of bias voltage

Figure 3.35 shows Shockley-Read generation (dotted line) and recombination (solid line). Carrier extraction relatively weakly influences Shockley-Read processes, and this is especially valid for generation. Naturally, the level of SR rates is strongly sensitive to the methods of photodetector single crystal growth. Thus in high-quality materials the levels of Shockley-Read g-r rates may be much lower than presented in Fig. 3.34. Their spatial distribution, however, qualitatively follows a similar functional dependence.

According to the calculated levels of g-r processes it is concluded that within the range of the chosen material parameters the prevailing process is remaining Auger generation. Total generation-recombination is primarily determined by this process.

3.6.6 *Current-Voltage Characteristics of Extraction-Exclusion Photodetector with Homojunction*

We considered current-voltage characteristics of extraction-exclusion photodiodes for the case of homojunction mercury cadmium telluride structures. All calculations were performed for the material composition x = 0.186. The dopant concentrations in the highly doped regions were 8×10^{17} cm^{-3} within the n$^+$ zone and 5×10^{18} cm^{-3} in the p$^+$ zone. A gradient junction was assumed, modeled by a tangent hyperbolic. The total width of the junctions was 0.1 µm each. The width of the p$^+$ region was 5 µm, and that of the n$^+$ region 0.5 µm.

Figure 3.36 shows a current-voltage characteristics of the described structure for an active region width of 2 µm, a doping level 3×10^{16} cm^{-3} within the range of operating temperatures 190–270 K. The expected increase of the maximum current with temperature is readily seen, as well as the increase of the necessary bias to reach that maximum. The negative dynamic resistance zone is marked for all of the temperatures from the investigated range. The ratio between the maximum and the minimum current is larger for higher operating temperatures.

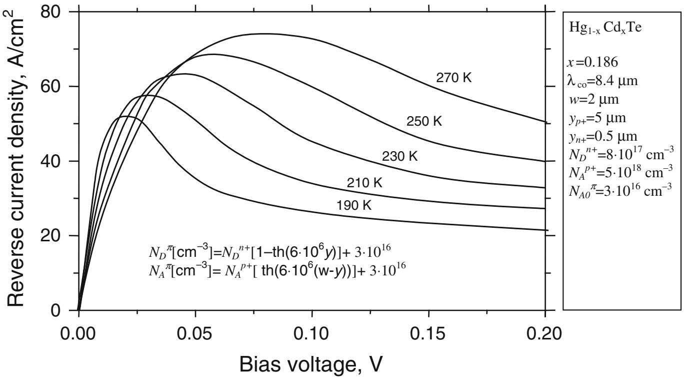


Fig. 3.36 Current-voltage characteristics of an HgCdTe $n^+\pi p^+$ type homojunction extraction diode for different operating temperatures

Figure 3.37 shows the current-voltage characteristics for the same structure, but for different doping levels of the active region at a temperature of 250 K. According to this diagram the current density in the negative dynamic resistance zone decreases with an increase of the p-doping level in the active region from the pseudoextrinsic value π via the extrinsic limit (about 5×10^{16} cm^{-3}) to p^- level (p^- is a notation for the doping level between π and p values). This means that for the given set of parameters we can choose the optimum doping level for the active region.

According to White [381] the results of the calculation for an n-type active region ($n^+\nu p^+$ structure) show that the obtained characteristics are always worse than for a corresponding $n^+\pi p^+$ diode, meaning that the carrier depletion is weaker, and the necessary current densities larger. According to the same author, in the

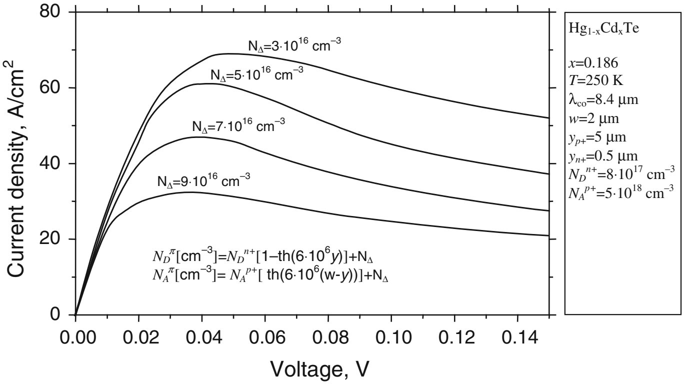


Fig. 3.37 Current-voltage characteristics of an HgCdTe $n^+\pi p^+$ homojunction extraction diode for different doping levels of the extraction zone

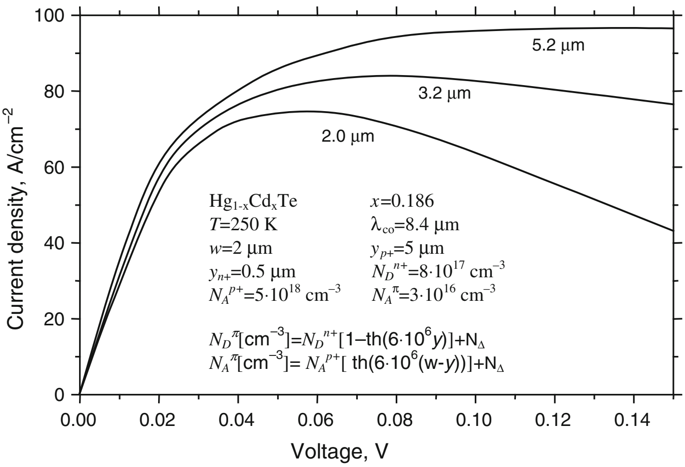


Fig. 3.38 Current-voltage characteristics of a homojunction HgCdTe extraction $n^+\pi p^+$ diode for various widths of the extraction zone

pseudointrinsic case the difference between the current density is about 5 %. In the extrinsic case it increases and finally a situation is reached where the negative resistance range is completely absent.

Figure 3.38 analyzes the influence of the thickness of the pseudoextrinsic zone to the current-voltage characteristics. The calculation was done for the standard set of parameters and an operating temperature of 250 K.

With an increase of the width of the active area the effect of the negative resistance becomes smaller and finally completely disappears. According to Sect. 3.5 the thickness of the pseudoextrinsic layer must be comparable to the diffusion length in the given material (in the considered case about 2.7 µm). If the thickness is much larger, the useful effect disappears altogether and the device behaves as two separate junctions in a series, one of them a hi-lo junction and the other an isotype junction.

3.6.7 Photoelectric Gain and Noise of Extraction Photodiode

Photoelectric gain in extraction diodes can be calculated in a manner analogous to that in exclusion structures, i.e., by dividing the detector along its length into infinitesimally thin elements dy and subsequently summing each of their contributions to the total detector current according to (1.93). The generation and recombination rates are given by (1.85) and (1.86), and the optical generation term is (1.80).

Figure 3.39 shows spatial distribution of photoelectric gain within a $p^+\pi n^+$ mercury cadmium extraction photodiode for different bias voltages. The detector parameters are given in the Fig. 3.39. Gain is given for the whole structure, including not only the depleted region, but also highly doped parts. An important

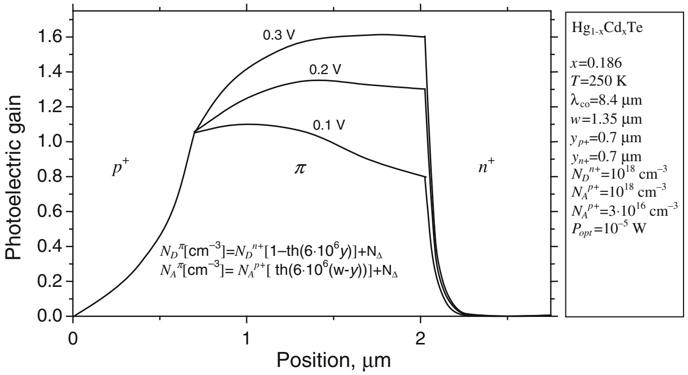


Fig. 3.39 Spatial distribution of photoelectric gain in a p$^+\pi$n$^+$ extraction diode for different bias levels

result is that in extraction diodes photoelectric gain reaches values higher than 1, i.e., a situation not typical for photovoltaic devices and more common for photoconductors.

Such result for extraction diodes was reported theoretically by Piotrowski et al. [382], and experimentally by Elliott et al. [362] and Skauli et al. [383]. Elliott et al. explained the appearance of Γ larger than 1 as a consequence of Auger generation (impact ionization), as well as that of the influence of redistribution of carriers within biased detector to the total conductivity of the device (the effect of mixed conductivity).

The next diagram, Fig. 3.40, shows spatial distribution of noise density factor (product of the sum of absolute values of generation and recombination rates and squared photoelectric gain, $(|G| + |R|)\Gamma^2$.

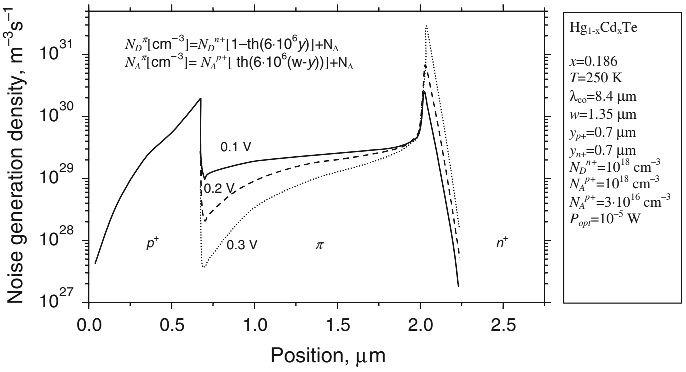


Fig. 3.40 Spatial distribution of noise generation density in an extraction-type diode for various values of bias

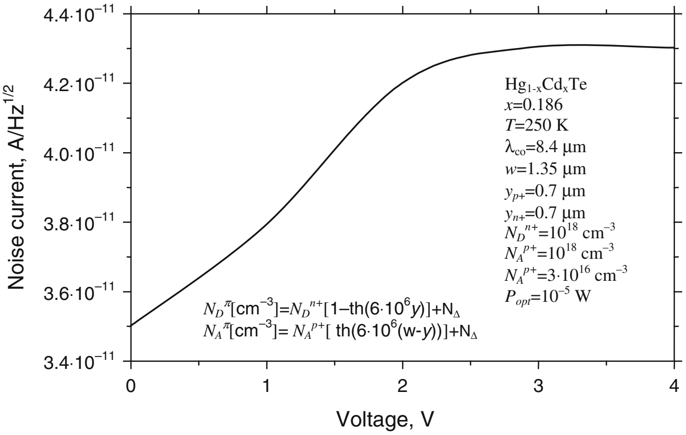


Fig. 3.41 Dependence of noise current on bias for a mercury cadmium telluride extraction diode

The influence of the n^+ zone to the total noise level in extraction devices is large, and bias increase even increases this influence. However, additional doping of this zone can be used to minimize the influence of g-r processes on the $n^+\pi$ junction.

Bias voltage increase decreases the noise component originating from the extraction zone. There is a very high influence of the p^+ zone to the total value of this noise component. The authors of this type of devices noted [384] that in a homojunction extraction diode the most part of the dark current and noise current originated from this zone and was generated by Auger seven processes. In the valence band the density of states is large, so that it would be very difficult to increase the doping level in p-material until degeneration is reached. Thus alternative solutions have been proposed: to use a wide-bandgap P-semiconductor instead of highly doped p-material, to use a layer which is at the same time highly doped and with a wider bandgap, or the fabrication of more complex structures with a larger number of highly doped buffer layers [362, 384]. All of these methods basically narrow the zone in the left part of the diagram, thus making it more similar to the right-hand side, and the consequence is that the effect of depletion of the extraction layer becomes much more marked.

Integration of the noise generation density along the whole length of the extraction diode gives the total noise current density for a given bias. Figure 3.41 shows the dependence of the total noise current for a given value of bias. Figure 3.41 shows the dependence of the total noise current on bias voltage for the case shown in Fig. 3.40. The values of the applied voltage were increased here up to 4 V, to observe the plateau of saturation in the characteristics of the frequency density of noise current which in this structure is reached at bias values above 2 V.

In the presented noise characteristics the low-frequency 1/f component was not taken into account, i.e., the presented dependences are valid only for high operating frequencies. Experiments show that the knee frequency of the 1/f noise in extraction devices reach values as high as several hundreds of megahertz [385].

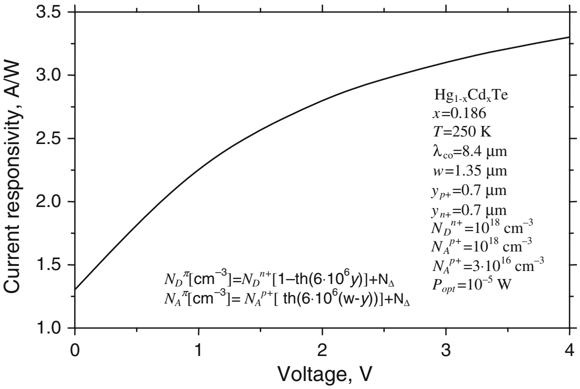
3.6.8 *Sensitivity and Detectivity of Extraction-Exclusion Detectors*

Figure 3.42 shows the dependence of current sensitivity of a homojunction mercury cadmium telluride $p^+\pi n^+$ extraction photodiode on bias voltage. The signal current was calculated as the difference between the dark current (obtained by van Roosbroeck-Shockley's model for a g-r term according to (1.85)–(1.86), i.e., without optical generation) and the current of illuminated detector (the optical term (1.80) is added to the g-r term for the dark current. The definition of sensitivity was used (signal current divided by the optical power of monochromatic excitation near the detector cutoff).

It can be seen that the improvement of sensitivity with bias for a homojunction diode is modest. The simplest and technologically easiest to implement extraction device is at the same time the least efficient.

The specific detectivity for the same diode is shown in Fig. 3.43. The same remark about the low efficiency of the structure is valid as above. However, even with such levels of improvement it can be seen that values of 4×10^8 cm Hz$^{1/2}$/W at 8.4 μm with cooling to 250 K can be reached. If extraction diodes with more complex profiles and with additional layers for performance improved are used, the values of specific detectivity under similar conditions exceed 10^9 cm Hz$^{1/2}$/W [386]. For the sake of comparison, the theoretical BLIP detectivity for a background at 300 K and radiating as blackbody is 6.24×10^{10} cm Hz$^{1/2}$/W. The conclusion is that the performance of extraction detector, in spite of all limitations, approaches the BLIP limit.

Fig. 3.42 Current sensitivity of a mercury cadmium telluride extraction homojunction diode versus bias voltage



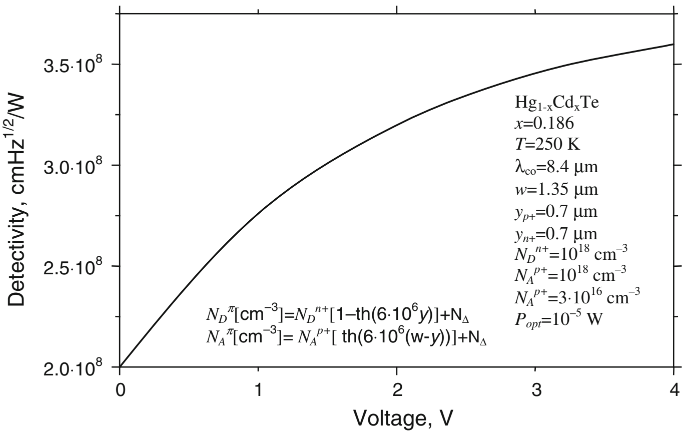


Fig. 3.43 Specific detectivity of a HgCdTe extraction diode versus applied bias

3.6.9 Experimental Results on Extraction-Exclusion Detectors

The first experimental works dedicated to the fabrication and characterization of Auger-suppressed extraction photodetectors handled mostly homojunction devices. Ashley et al. [363] fabricated three-layer $p^+\pi n^+$ structures in $In_{1-x}Al_xSb$. They utilized molecular beam (MBE) and (100)-oriented InSb substrate. The doping level of the π zone was about 10^{15} cm^{-3}, its thickness 0.75–3 μm. The dopant concentrations in highly doped zones were higher than 10^{18} cm^{-3}. Etching was used to fabricate mesa structures with an area of about 10^{-4} cm^2 and passivation with anodic oxide was used.

These first extraction diodes had several shortcomings. As the authors themselves noted in [363], leakage currents were excessive, which was connected with the nature of the p^+ region. In addition to that, anodic oxide had a shunting effect to the active area. The same author subsequently did experiments with four-layer $p^+P^+\pi n^+$ structures [363] in $In_{1-x}Al_xSb$ where P^+ layer was very thin and strained. Leakage currents were significantly decreased and the devices showed at room temperature a performance unattainable with homojunction devices at temperatures above 250 K.

A homojunction solution utilized additional guard rings and the proximity effect (Fig. 3.28a). This solution was patented by Elliott et al. [376]. At an operating temperature of 200 K and with a cutoff λ_{co} = 9.3 μm these diodes reached 10^9 cm Hz$^{1/2}$/W.

Elliott et al. started experimental fabrication of PπN-type extraction diodes, as well as $P\pi n^+$ and $P\nu n^+$ [361, 362]. They used HgCdTe which enabled them to fabricate heterostructures without built-in stress. They used metalorganic vapor phase epitaxy (MOVPE) and IMP procedure (fabrication of interdiffusion multilayers). The substrate was GaAs or CdZnTe (4 % Zn), and after deposition the samples were annealed in an open system to remove Hg vacancies. The bandgap of the wide-bandgap layers was typically 2.5–3 times larger than that in the π region.

Arrays of 64 HgCdTe elements with mesa diodes (20–80) μm were fabricated [362]. The doping level in the π region was 2–4 × 10^{18} cm^{-3}, the region width 5 μm. The n-type dopant was iodine, p-type dopant was arsenic. In these devices an excessive 1/f noise component was noted with a knee at about 10 MHz. The authors noted that the diodes they fabricated had a photoelectric gain $\Gamma > 1$.

The following fit is valid for current-voltage characteristics of these diodes

$$I_{\max}/I_{\min} = -4.02 + 0.03729T - 5.92 \cdot 10^{-5}T^2 \tag{3.115}$$

i.e., the ratio of the maximum and minimum current increases with temperature. At near-room temperature the empirical relation is valid

$$J_{\min} = 13,000e^{-qE_g/k_bT}. \tag{3.116}$$

MOVPE was used to fabricate $In_{1-x}Al_xSb$ $p^+P^+\pi n^+$ type extraction devices [365]. Dopants and the structures were similar to those of the mentioned HgCdTe diodes.

Heterodyne detection was investigated for PπN detectors [386]. At 40 MHz a current sensitivity of 2.9 A/W and NEP of about 2 × 10^{-9} $\text{W/Hz}^{1/2}$ were measured.

A large number of experiments with extraction devices was dedicated to negative luminescence. Early results include InAlSb (wavelength range 5.5–5.8 μm) [365] and HgCdTe devices (range 5–7 μm) [364]. Nash et al. [387] presented a mercury cadmium telluride negative luminescence device with a luminescence peak at 7.6 μm at 296 K with an equivalent T reduction of 43 K due to Auger process suppression.

In later generation devices exclusively heterostructures have been used. Skauli et al. [383] reported MBE-fabricated $Hg_{1-x}Cd_xTe$ extraction diodes using silver as dopant in P region and indium in the N region. All regions in their three-layer structures were about 3.5 μm thick, and they worked at $\lambda_{co} = 9$ μm for a temperature of 295 K. These authors also noticed photoelectric gain higher than zero within the active areas of their diodes.

Numerous works were dedicated to investigation of 1/f noise of extraction devices [243]. A numerical model also was presented for the calculation of 1/f noise in mid-wave infrared HgCdTe detectors with arbitrarily nonuniform doping profiles and band structures [388, 389]. This component of noise is probably the principal hindrance to a more widespread use of nonequilibrium long-wavelength photodetectors. It has been established that 1/f noise increases with dynamic resistance, reaches its maximum in the region of negative resistance and becomes flat around the minimum of the current-voltage characteristics. It varies significantly from device to device, both regarding its magnitude and the knee frequency.

At wavelengths about 10.6 μm and at temperatures of 295 K the knee frequency reached 100 MHz [385]. $In_{1-x}Al_xSb$ and $Hg_{1-x}Cd_xTe$ detectors were investigated and in both types a strong correlation was established with their dark current. At temperatures below 100 K only g-r component appeared, at about 140 K noise with Lorentzian spectrum appeared, typical for excess noise due to g-r processes at single traps [385], while at near-room temperatures 1/f noise became dominant. The authors concluded that 1/f noise at elevated temperatures appears as a consequence of a large number of uncorrelated Lorentzian spectra. This could mean that the main cause of excessive 1/f noise is surface states or surface tunneling, i.e., the dislocation density in bulk and at the junctions is not the decisive factor.

Schaake et al. measured HOT (higher operating temperature) mercury cadmium telluride photodiodes with a cutoff at 5 μm at a temperature of 250 K [390]. They used an arsenic-doped high density vertically integrated diode fabricated by liquid phase epitaxy.

3.6.10 Advantages and Disadvantages of Extraction-Exclusion Devices

Compared to other types of Auger suppressed photodetectors, extraction devices have the lowest dark currents. This means a number of their advantages over other types. Their operation is far from the hot carrier region, the level of g-r processes is lower, the problems with the structure heating are less pronounced. These devices are able to operate at relatively largest temperatures.

Extraction diodes are the most advanced of all nonequilibrium detectors. The fabrication of these detectors is fully compatible with standard narrow-bandgap semiconductor technologies and with standard detector circuitry. Their function does not cause interference with surrounding circuitry, as is the case with magnetoconcentration devices.

A problem is the relative complexity of epitaxial technologies procedures required to achieve satisfactory performance. It is necessary at the same time to have a very accurate composition control, good matching of the crystal lattice, doping level control, and relatively low doping level of the active region (although the latter condition is much less stringent than in the case of exclusion devices).

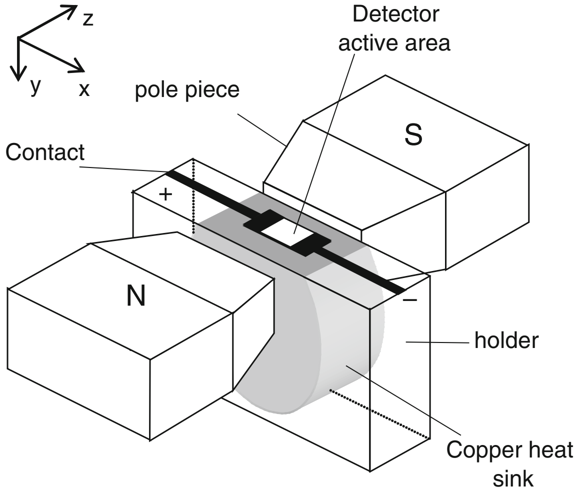
The most serious problem with extraction detectors at this time is their 1/f noise. The knee frequency of these devices is very high and reaches 20 MHz to 100 MHz. Large efforts have been directed to its investigation [391, 392], but currently the topic still remains an open issue.

3.7 Galvanomagnetic Methods—Magnetoconcentration Photodetector

3.7.1 Physical Concept

Magnetoconcentration or Suhl effect (Harry Suhl) [393] utilizes crossed electric and magnetic fields directed in parallel to the surface of a single crystalline semiconductor slab (here: the photodetector), as shown in Fig. 3.44. The presented

Fig. 3.44 Schematic presentation of magnetoconcentration detector



geometry is similar to that of a typical photoelectromagnetic (PEM) detector. The slab thickness must be small enough to disable carrier motion along a cycloid (an equivalent requirement is that the magnetic field is low enough). A variation of the presented scheme is to use separate contacts for biasing and signal readout.

Magnetoconcentration detectors could straightforwardly benefit from the use of microfabrication technologies, since this would mean smaller permanent magnets and thus more compact devices, but also fields at the same time higher and better confined.

If the Lorentz force is oriented from the active area, charge carriers are pushed into the bulk. If their free path is larger than the slab thickness, most of the carrier reaches the bottom surface and accumulates there. Thus magnetoconcentration is a change of spatial distribution of carriers caused by crossed electric and magnetic field.

A possible method to utilize magnetoconcentration to suppress Auger generation-recombination is to process the surface where carriers accumulate to obtain high recombination rate ($S_2 \to \infty$), while keeping it low at the active surface ($S_1 \approx 0$). A method typically used to tailor the surface recombination rate in magnetoconcentration devices is chemomechanical polishing. Alternatively, one may utilize MEMS/NEMS approach to fabricate ordered micro or nanostructured 2D arrays of pillars or holes on the surfaces, thus achieving higher repeatability.

At the surface S_2 the carrier concentration remains very close to the equilibrium one. Since the generation rate in the rest of the slab (including the opposite surface) is low, it cannot compensate carrier concentration decrease caused by recombination at the S_2 surface and thus the overall concentration in the sample decreases, its values being lower for the positions closer to the surface with S_1. The procedure corresponds to exclusion, but in this case both carrier types are deflected.

If the direction of electric or magnetic field is opposite (carriers are deflected to the low recombination surface) accumulation of carriers occurs.

The limit of concentration decrease of the majority carriers in the exclusion case is dopant concentration. In order to preserve electroneutrality, the concentration of minority carrier decreases several orders of magnitude more. The result is an intensive carrier depletion near the top surface and thus Auger suppression.

The level of suppression will depend on thermal generation rate and the intensity of applied field. It is usually said that magnetoconcentration plays a significant role at high fields and large injections [394].

Similar to all other nonequilibrium devices, Auger suppression is significant only in starting concentrations of electrons and holes are comparable (near-intrinsic material), and negligible in strongly doped semiconductor.

3.7.2 *History of Magnetoconcentration Effect Research*

Magnetoconcentration (Suhl effect) was first mentioned in 1949 in connection with point-contact transistor where base resistance was modulated by external magnetic field [393]. Negative luminescence was observed in magnetoconcentration experiments of Ivanov-Omski et al. [338]. At the beginning of 1980s a theoretical explanation was given for negative luminescence at narrow-bandgap direct materials like InSb and HgCdTe [395, 396]. Japanese authors Morimoto and Chiba [331] investigated the "magnetoelectric effect", i.e., the possibility to use magnetoconcentration for lasing and analyzed negative luminescence caused by Lorentz force calling it "induced absorption".

A biased photoconductor under the influence of magnetoconcentration for the case of low magnetic fields was investigated by Malyutenko and Teslenko [397, 398]. They denoted the effects of Lorentz force to carriers in such a device the "transversal sweep-out" ("a semiconductor with induced anisotropy of conductivity"). Their work included both theoretical and experimental investigation, and the magnetic induction they utilized did not exceed 0.2 T.

In 1990 Djurić and Piotrowski proposed and theoretically elaborated the use of magnetoconcentration for nonequilibrium Auger suppression [67, 334]. They called their structures EMCD (Electro-Magnetically Carrier Depleted). Their further works analyze EMCD in InAsSb and HgCdTe in some depth [243]. The paper [67] proposed the use of magnetoconcentration to fabricate FET devices with Auger suppression. In [399] some problems with magnetoconcentration experiments are considered. Jakšić and Djurić [400] consider the time response of nonequilibrium magnetoconcentration photodetectors.

3.7.3 Model of Nonequilibrium Magnetoconcentration Detector

A magnetoconcentration detector is described by the set of equations (3.57) to (3.64). The continuity equations have the identical form to (3.65) and (3.66). Current density in the direction of the magnetic field is given by (3.59). Together with the electroneutrality equation (3.63) and (3.64) one obtains a second order differential equation for the minority carrier concentration. For n-type semiconductor this equation is

$$\frac{\partial p}{\partial t} = C_1(p)\frac{\partial^2 p}{\partial y^2} + C_2(p)\left(\frac{\partial p}{\partial y}\right)^2 + C_3(p)\frac{\partial p}{\partial y} - g. \tag{3.117}$$

where the coefficients C_1, C_2, and C_3 are

$$C_1 = -\frac{\mu_n\mu_p k_b T}{\gamma_{eb}\tau_{po}}\left\{\frac{p}{2}\left(\tau_{no} - \tau_{po}\left(1 + \frac{p_{ni}}{p_{nb}}\right) - n\tau_{po}\right)\right\}l \tag{3.118}$$

$$\begin{aligned} C_2 = {} & \frac{\mu_n\mu_p k_b T}{\gamma_{eb}^2}\frac{p\gamma_{eb}\left\{N_D\tau_{n0}\tau_{p0} + \left(\tau_{n0} + \tau_{p0}\right)\left\lfloor(p + p_1)\tau_{n0} + (n + n_1)\tau_{p0}\right\rfloor\right\}}{\left[(p + p_1)\tau_{n0} + (2n - p + N_A + n_1)\right]^2} \\ & + \left\{\frac{\mu_{eb}}{2}\left[\tau_{n0} + \tau_{p0}\left(1 + \frac{p_{ni}}{p_{nb}}\right)\right] - \mu_{hb}\tau_{p0}\right\}\left\{\frac{p}{2}\left[\tau_{n0} + \tau_{p0}\left(1 + \frac{p_{ni}}{p_{nb}}\right)\right] - n\tau_{po}\right\} \end{aligned} \tag{3.119}$$

$$C_3 = \frac{q\mu_n\mu_p\left(\mu_n + \mu_p\right)E_x B}{\gamma_{eb}\tau_{p0}}\left\{\frac{\mu_{hb}p^2}{2}\left[\tau_{n0} - \tau_{p0}\left(1 + \frac{p_{ni}}{p_{nb}}\right)\right] - \mu_{eb}n^2\tau_{p0}\right\} \tag{3.120}$$

and the following notation is used

$$\mu_{eb} = \mu_n\left(1 + \mu_p^2 B^2\right) \tag{3.121}$$

$$\mu_{nb} = \mu_p\left(1 + \mu_n^2 B^2\right) \tag{3.122}$$

$$\gamma_{eb} = \mu_{eb}n + \mu_{hb}p \tag{3.123}$$

$$p_{ni} = \frac{2N_D\tau_{n0}\tau_{p0} + \left(\tau_{n0} + \tau_{p0}\right)}{\tau_{n0}^2}\left[p_1\tau_{n0} + (n_1 - N_A)\tau_{p0} + p\left(\tau_{n0} + \tau_{p0}\right)\right] \tag{3.124}$$

$$p_{nb} = \sqrt{4N_D\left[n_1 + p\frac{\tau_{n0}}{\tau_{p0}}\right] + \left[\frac{p_1\tau_{n0} + (n_1 - N_A)\tau_{p0} + p\left(\tau_{p0} + \tau_{n0}\right)}{\tau_{p0}^2}\right]}. \tag{3.125}$$

The electron concentration n according to the electroneutrality equation and utilizing the above notation is

$$n = \frac{1}{2}p\left(1 - \frac{\tau_{no}}{\tau_{po}}\right) - \frac{1}{2}\left(p_1\frac{\tau_{no}}{\tau_{po}} + n_1 + N_A\right) + p_{nb}. \tag{3.126}$$

The boundary conditions for the carrier transport equation in the magnetoconcentration detector are posed for the top and bottom surface, assuming that their surface recombination rates are described by the Shockley-Read model, i.e., that $S = 1/\tau_{no} = 1/\tau_{po}$. The second assumption often met in literature (e.g., [401]) is that the trap distribution function at the surface is always of Maxwell-Boltzmann type (nondegenerate semiconductor). In that case Shockley-Read expression for recombination rate (1.70) becomes

$$R_{SR} = \frac{p(y)n(y) - n_0p_0}{n + p + 2n_i ch\left(\frac{E_t - E_i}{k_bT}\right)}. \tag{3.127}$$

where the index "0" denotes the equilibrium concentration. If $E_t = E_i$, the above expression reduces to

$$R_{SR} = \frac{p(y)n(y) - n_0p_0}{n + p + 2n_i}. \tag{3.128}$$

For a low surface recombination rate (the top surface, $S = S_1$) the concentrations are $n(y = 0) \ll n_i$, $p(y = 0) \ll n_i$. The opposite is the case for high recombination rates ($S = S_2$). Thus

$$\begin{aligned} \frac{1}{q}J &= S_1\frac{p(y)n(y) - n_0p_0}{n_0 + p_0}, \quad y = 0 \\ \frac{1}{q}J &= S_2\frac{n_0p_0 - p(y)n(y)}{n + p}, \quad y = d. \end{aligned} \tag{3.129}$$

For the case of totally ionized donors (majority dopants) the current density can be explicitly represented as a function of the concentration of one type of carriers (holes for the case of the n-type semiconductor)

$$J_y = a(p) + b(p)\frac{dp}{dy}, \tag{3.130}$$

where the coefficients a and b are

$$a(p) = \frac{q\mu_n\mu_p\left(\mu_n + \mu_p\right)(p - N_A)p\,BE_x}{\mu_n(p - N_A)\left[1 + \left(\mu_p B\right)^2\right] + \mu_p p\left[1 + (\mu_n B)^2\right]}, \tag{3.131}$$

$$b(p) = \frac{\mu_n D_p(p - N_A) + \mu_p D_n p}{\mu_n(p - N_A)\left[1 + \left(\mu_p B\right)^2 + \mu_p p\left[1 + (\mu_n B)^2\right]\right]}. \tag{3.132}$$

The differential equation describing the concentration of minority carriers has the same form as (3.117), but the coefficients C$_1$, C$_2$, and C$_3$ are now

$$C_1 = \frac{q\mu_n\mu_p}{kT}\frac{2p - N_A}{(p - N_A)\mu_{eb} + p\mu_{hb}}, \tag{3.133}$$

$$C_2 = \frac{N_A(\mu_{hb} - \mu_{eb})}{(2p - N_A)[(p - N_A)\mu_{eb} + p\mu_{hb}]}, \tag{3.134}$$

$$C_3 = \frac{q}{kT}(\mu_n + \mu_h)E_x B\frac{\mu_{eb}(p - N_A)^2 + \mu_{hb}p^2}{(2p - N_A)[(p - N_A)\mu_{eb} + p\mu_{hb}]}, \tag{3.135}$$

where μ_{eb} and μ_{hb} are given by (3.121) and (3.122).

The boundary conditions are slightly modified

$$\frac{1}{q}J = S_1\frac{p(0)(p(0) - N_A) - n_0 p_0}{n_0 + p_0}, \tag{3.136}$$

$$\frac{1}{q}J = S_2\frac{n_0 p_0 - (p(d) - N_A)}{2p(d) - N_A}. \tag{3.137}$$

3.7.4 Analytical Consideration of Nonequilibrium Magnetoconcentration Photodetector

We consider a ν-type magnetoconcentration detector. The crossed electric and magnetic field cause the depletion of a part of its volume. In an ideal case this would mean that the Auger g-r processes would be sufficiently suppressed to be negligible in comparison to radiative recombination. We further assume that material is sufficiently pure to allow to neglect Shockley-Read processes in bulk. Thus the generation term in the continuity equation reduces to the radiative term. Besides the bulk radiative lifetime we include in it a term of the form 1/*S* (*S* is surface recombination rate which is of SR nature). Then we assume that both of these terms have the identical dependence on the carrier concentration, so that the total g-r rate can be expressed as $n_i/2\tau_{\mathrm{ri}}$ (here τ_{ri} is the effective radiative lifetime which includes not only the real radiative lifetime with reabsorption but also the surface recombination rate).

The n-type sample is pseudo-extrinsic and thus $n \gg p$; we also know that $\mu_{\mathrm{n}} \gg \mu_{\mathrm{p}}$ and $D_n \gg D_p$, and since the active area is depleted, due to the effects of the space charge we will have $\mathrm{d}p/\mathrm{d}y \approx \mathrm{d}n/\mathrm{d}y$.

We start from the expression for the current density in crossed electric and magnetic fields (3.59). In the depleted part the carrier concentration varies slowly with position, thus we may assume that carrier transport is mainly caused by drift. This means that

$$J_{\mathrm{tot}} \approx \frac{q\mu_n\mu_p\left(\mu_p + \mu_n\right)np}{\mu_n n\left(1 + \mu_p^2 B^2\right) + \mu_p p\left(1 + \mu_n^2 B^2\right)} BE_x. \tag{3.138}$$

Here $\mu_{\mathrm{p}}B \ll 1$ since hole mobility is of the order of 0.01 m^2/Vs and B is max. 2 T. On the other hand, μ_{n} is of the order 1 m^2/Vs and thus the value of the right term in parenthesis is between 1 and 10. Thus, the right term can be neglected, since $\mu_{\mathrm{p}}p$ is several orders of magnitude smaller than $\mu_{\mathrm{n}}n$. The value of the term in remaining parenthesis is approximately μ_{n}. If we introduce the notation $N_\Delta = |N_A - N_D|$, we obtain

$$J_{\mathrm{tot}} = q\mu_n\mu_p pBE_x = q\mu_n\mu_p BE_x(n - N_\Delta), \tag{3.139}$$

since in the depleted region $p = n - N_\Delta$. Then

$$J_{\mathrm{tot}} = \begin{cases} q\mu_n\mu_p BE_x(n - N_\Delta) & \text{n-type} \\ \dfrac{q\mu_n^2 n}{1 + \mu_n^2 B^2} E_x B & \text{p-type,} \end{cases} \tag{3.140}$$

According to the continuity equation for electrons (3.57) and the expression (3.140) for n-type material we obtain

$$\frac{1}{q}\frac{\mathrm{d}}{\mathrm{d}y}\left[q\mu_n\mu_p BE_x(n - N_\Delta)\right] = -g_{\mathrm{rad}}, \tag{3.141}$$

thus

$$\frac{dn}{dy} = -\frac{g_{\mathrm{rad}}}{\mu_n\mu_p BE_x}, \tag{3.142}$$

and after integration, denoting the electron concentration at the low recombination surface as n_0, we obtain for n-type semiconductor

$$n = -\frac{g_{\mathrm{rad}}}{\mu_n \mu_p B E_x} y + n_0. \tag{3.143}$$

An equivalent procedure gives for p-type material

$$n = -\frac{g_{rad}\left(1 + \mu_n^2 B^2\right)}{\mu_n^2 B E_x} y + n_0, \tag{3.144}$$

thus the spatial dependence of carrier concentration is approximately linear until the point is reached where the approximation no longer holds.

For the sake of convenience we set our coordinate system to $n_0 = 0$. By directly replacing (3.143) into (3.140) the following is obtained for n-type sample

$$J_{\mathrm{tot}} = -q g_{\mathrm{rad}} y - q \mu_n \mu_p B E_x N_\Delta. \tag{3.145}$$

while the replacement of (3.144) into (3.140) gives for p-type material

$$J_{\mathrm{tot}} = -q g_{\mathrm{rad}} y. \tag{3.146}$$

Thus, in both cases the current carrier linearly changes with spatial coordinate until the point is reached where the approximation is no longer valid.

The thickness of the depleted region is obtained using the expression for the diffusion component of current. Since depletion means that drift processes prevail, this means that the limit of the depleted region will be in the point where diffusion is no longer negligible, i.e., where it exceeds a certain percentage of total current density. If we denote that percent as r_{mc}, it will be valid

$$r_{\mathrm{mc}} J_{\mathrm{tot\,g}} = J_{\mathrm{difuz}}, \tag{3.147}$$

where J_{totg} is total current density at the boundary of the depleted region, and J_{difuz} is diffusion current density.

An arbitrary adopted value for r_{mc} is 0.03.

The approximate value of the diffusion current is obtained starting from the diffusion term in presence of magnetic field

$$J_{\mathrm{dif}} = \frac{\mu_n D_p n \frac{dp}{dy} + \mu_p D_n p \frac{dn}{dy}}{\mu_n n\left(1 + \mu_p^2 B^2\right) + \mu_p p\left(1 + \mu_n^2 B^2\right)}. \tag{3.148}$$

According to the consideration of (3.139), for n-type material the denominator in (3.148) is approximately $\mu_n n$, and for p-type material $\mu_p p(1 + \mu_n^2 B^2)$. On the other hand, the derivatives of p and n are equal. Since $\mu_n D_p n \gg D_n p$, it means that

$$J_{\mathrm{dif}} \approx D_p \frac{dn}{dy} = -D_p \frac{g_{\mathrm{rad}}}{\mu_n \mu_p B E_x}. \tag{3.149}$$

Replacing the above equation into (3.145) and using the expression for the total current density (3.140) while y is equal to the critical thickness of the depleted region L_{krit} we obtain for n-type material

$$L_{\mathrm{krit}} = d - \frac{D_p}{q r_{\mathrm{mc}} \mu_n \mu_p B E_x} + \frac{\mu_n \mu_p B E_x N_\Delta}{g_{\mathrm{rad}}}, \tag{3.150}$$

and for p-type material

$$L_{\mathrm{krit}} = d - \frac{D_n}{q r_{\mathrm{mc}} \mu_n^2 B E_x}. \tag{3.151}$$

Expressions (3.150) and (3.151) can be used to estimate the depleted region thickness for different applied fields and varying detector parameters.

Figure 3.45 shows the critical length and concentration profile calculated by the approximate analytical model, as well as their values numerically calculated according to the full model (3.57)–(3.64), i.e., (3.117) do (3.129).

Figure 3.45a shows the dependence of the critical thickness of magnetoconcentration device on applied voltage for different magnetic inductions. It can be seen that there is a threshold voltage below which there is no depletion of the active region. If bias exceeds this value, the thickness of the depleted region sharply increases and then finally reaches saturation. The threshold voltage increases with magnetic field decrease.

For a given value of magnetic field and at a constant temperature the product of the applied electric field and the thickness of the part which is not depleted (i.e., the difference between the total thickness and the critical thickness) approximately remains constant. It has the dimensions of voltage and it is here denoted as the critical voltage, U_{crit}. For the case shown in Fig. 3.45a its value is $U_{crit} \approx 10$ mV.

For this calculation the real experimental values were taken for the doping level which according to [50] are necessary to reach nonequilibrium suppression in the active region of the detector.

Figure 3.45b shows the spatial dependence of carrier concentration across the EMCD device. A comparison with accurate model is also shown. It can be seen that at a certain distance from the top surface (right side of the diagram corresponds to the top surface) the numerically calculated concentration (dashed) becomes significantly larger than the analytical approximation (solid). This distance is equal to the critical thickness.

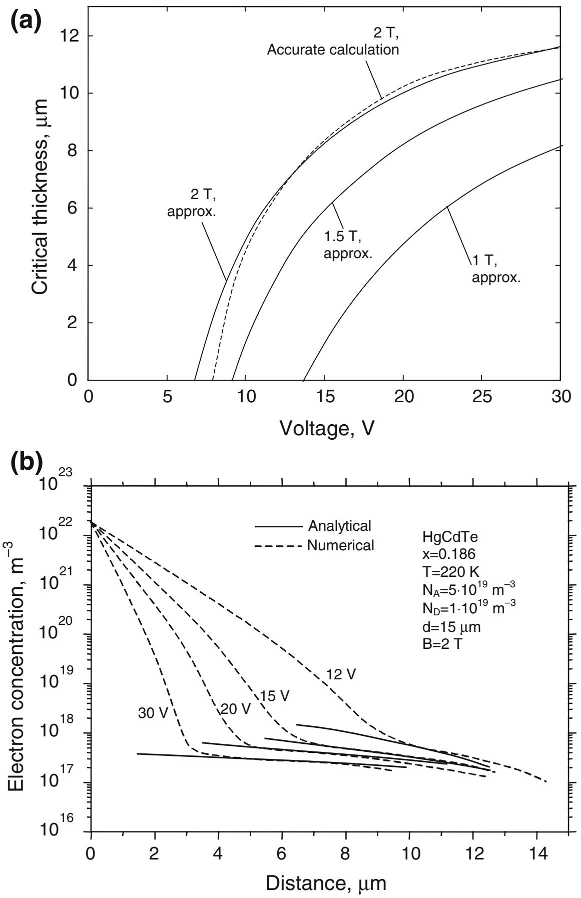


Fig. 3.45 Results of approximate analytical model of magnetoconcentration detector and their comparison with numerical simulation according to full mode. **a** Critical thickness versus bias for a magnetoconcentration HgCdTe p-type photodetector for different values of magnetic induction; T = 220 K, x = 0.187, d = 15 μm, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 10^{19}$ m^{-3}. Dashed line numerical calculation using accurate model; full lines analytical approximation. **b** Spatial dependence of carrier concentration across the magnetoconcentration EMCD device (y-axis)

3.7.5 Carrier Distribution in Nonequilibrium Magnetoconcentration Detector

Figure 3.46 shows the distribution of minority carriers in p-type nonequilibrium magnetoconcentration photodetector for different values of bias voltage. Larger values of bias and thicker detector slabs mean stronger depletion (several orders of magnitude).

There are two distinct zone in the concentration profile, denoted in Fig. 3.46 as A and B. Each bias or detector thickness increase in the zone A results in a dramatic depletion, while in the B zone there is no significant further concentration decrease. The boundary line (dotted line) corresponds to the critical thickness of depleted region L_{krit}, approximately given by (3.151).

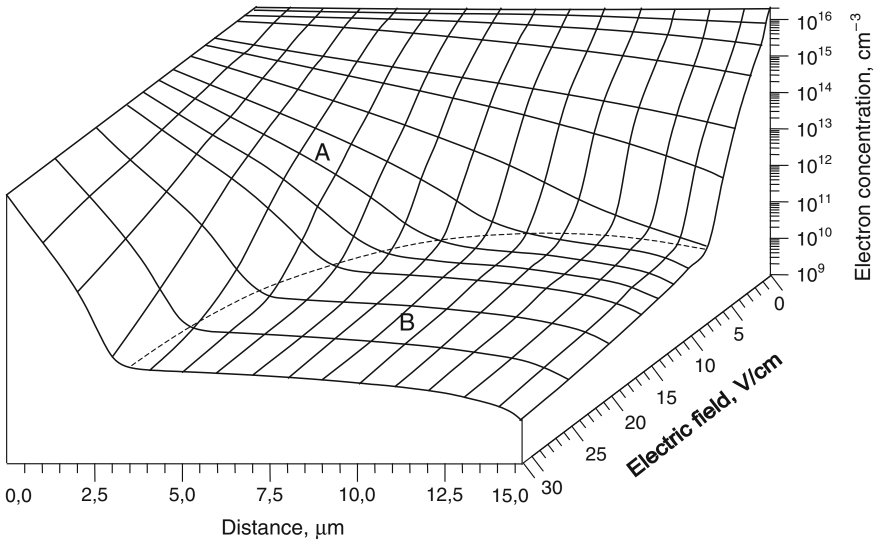


Fig. 3.46 Spatial distribution of electron concentration in p-type nonequilibrium magnetoconcentration photodetector. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ µm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ µm

It can be seen that the physical meaning of the critical thickness is that it is actually the magnetoconcentration exclusion length. On one side, it is insufficient to bias the detector without achieving a depletion beyond the boundary curve. On the other hand, there is no sense in increasing bias to obtain critical thickness much larger than the absorption length within the active region. Each excessive bias increase only heats carriers and does not contribute performance improvement (i.e., does not suppress Auger processes).

A detector with a suboptimal critical thickness will have insufficient Auger suppression and decreased absorption, i.e., excessive noise levels and suboptimal quantum efficiency. The one with excessive critical thickness will in best case have performance practically identical to the optimal case, and in worst case the excessive bias will cause undesired oscillations and other hot carrier effects without contributing the detector performance.

An equivalent consideration is valid for the case when the electric field is constant and magnetic induction is changed.

Figure 3.47 shows hole concentration in the same detector as in Fig. 3.46. It is seen that it drops to the impurity level, while electron concentration is several orders of magnitude lower. The optimum bias-thickness ratio curve can be seen in this dependence too.

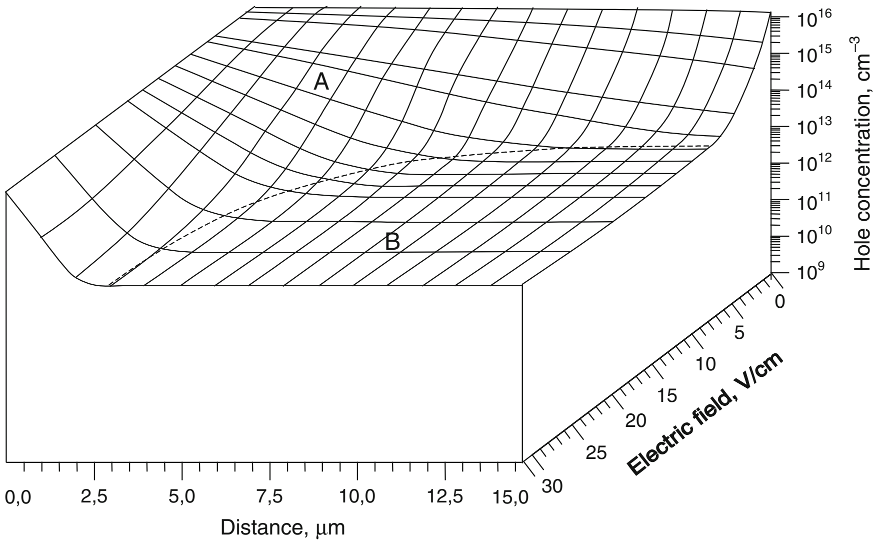


Fig. 3.47 Spatial distribution of hole concentration in p-type nonequilibrium magnetoconcentration photodetector. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ μm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ μm

3.7.6 Recombination Rates in Nonequilibrium Magnetoconcentration Photodetector

Figure 3.48 shows the dependence of the Auger processes rate on bias across a magnetoconcentration detector. Obviously, the recombination rate drop with bias is much larger than the generation rate decrease. For the set of parameters chosen for Fig. 3.48 the suppression of generation rate is three orders of magnitude, but recombination rate is suppressed seven orders of magnitude more. The same like carrier concentration, Auger generation, and recombination rate become saturated when the product of applied field and the detector thickness exceed the critical value.

Figure 3.49 shows the dependence of radiative recombination on bias along a magnetoconcentration photodetector without taking into account reabsorption effects. Figure 3.50 shows Shockley-Read generation (dotted) and recombination rate (solid). As expected, the influence of nonequilibrium suppression is much weaker than in the case of Auger processes, although the dependencies in both cases qualitatively follow the same shape.

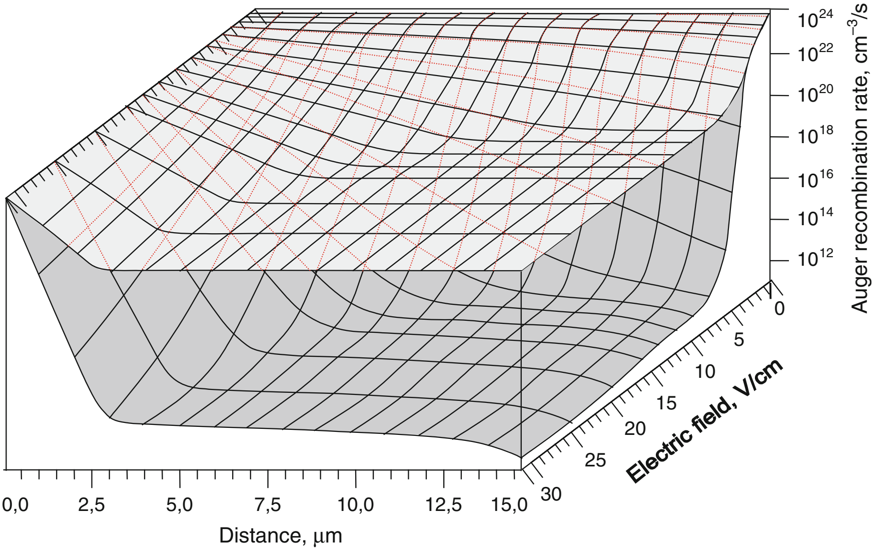


Fig. 3.48 Auger generation and recombination rates across a magnetoconcentration $Hg_{1-x}Cd_xTe$ detector for different values of bias. Top surface generation rate, bottom surface recombination rate. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ µm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ µm

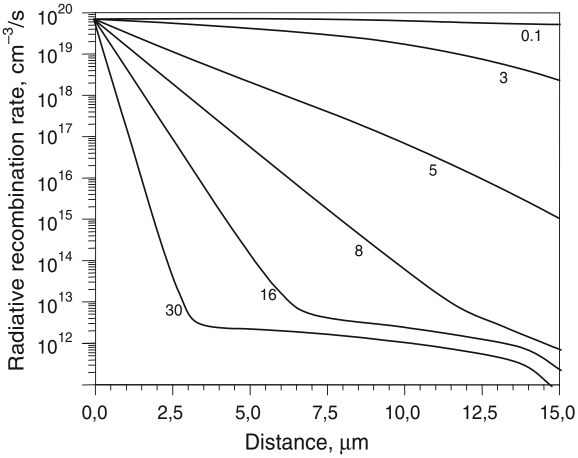


Fig. 3.49 Radiative recombination rate across a magnetoconcentration $Hg_{1-x}Cd_xTe$ photodetectors for different values of bias. Radiative generation rate is independent on position and is equal to 7.124×10^{19} cm^{-3}/s. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ µm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ µm

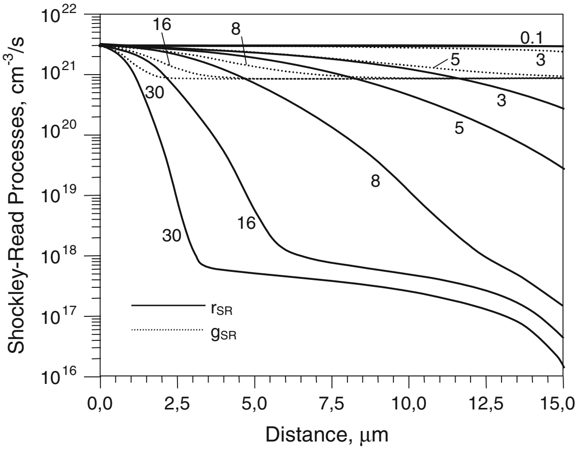


Fig. 3.50 Shockley-Read process rates across a magnetoconcentration $Hg_{1-x}Cd_xTe$ photodetector for different values of bias voltage. Solid recombination, dotted generation. Numbers near each curve denote bias voltage in volts. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ μm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ μm

3.7.7 Current-Voltage Characteristics of Nonequilibrium Magnetoconcentration Photodetectors

Figure 3.51 shows the dark current profile across the magnetoconcentration photodetector for different operating temperatures. As expected, the rise of the operating temperature increases dark current. Figure 3.52 shows the same dependence for different values of bias voltage at 220 K.

Figure 3.53 shows detector signal current (the difference between the current of the illuminated detector and the dark current). The signal current reaches its maximum near the high recombination rate surface and its dependence on position is nearly linear for the most part of the detector.

Fig. 3.51 Spatial distribution of dark current of a magnetoconcentration detector for different operating temperatures

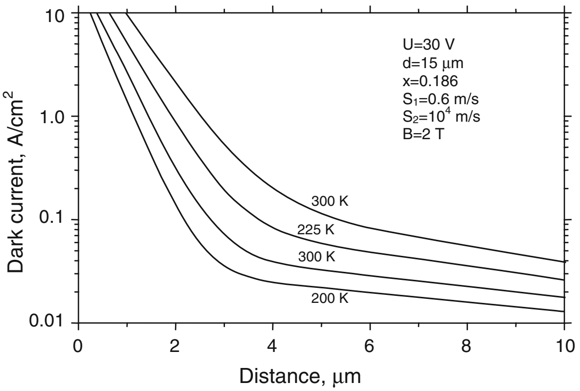


Fig. 3.52 Spatial distribution of dark current of a magnetoconcentration detector for different values of bias voltage

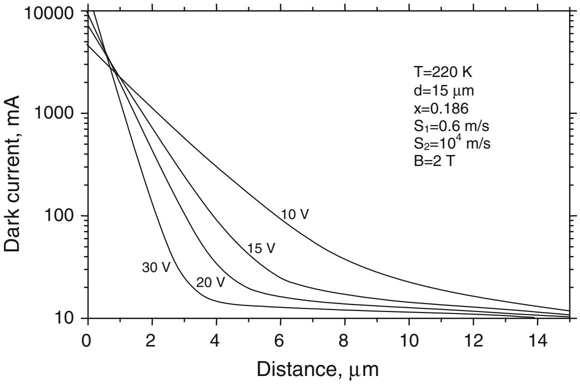
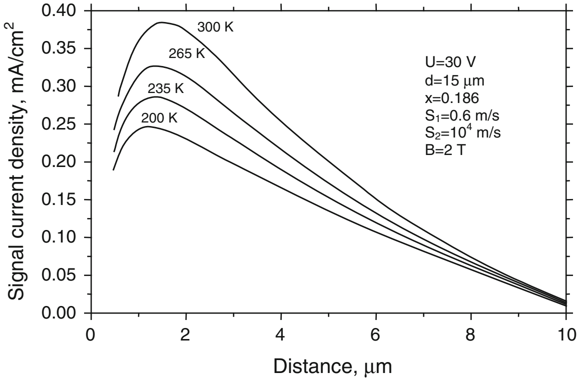


Fig. 3.53 Spatial distribution profile for signal current across HgCdTe magnetoconcentration detector



The total current of a magnetoconcentration photodetector in x direction is obtained as the integral of the expression for $J(y)$ over all values of y:

$$I/w = \int_0^d J(y)dy, \tag{3.152}$$

where the above expression denotes the current density per unit length (current across the width of the detector).

Assume that the external electric field may be approximately expressed as a ratio between the bias voltage and the sample length, $E = U/l$. This is valid if the electric field dependence on the position x is linear, which is satisfied for sufficiently depleted photodetector. In that case the current-voltage characteristics of the magnetoconcentration photodetector is given as

$$I = qw\frac{U}{L}\int_0^d \left(\mu_p p + \mu_n n\right)dy - B\left(\mu_n + \mu_p\right)\int_0^d J_y dy. \tag{3.153}$$

Fig. 3.54 Calculated UI characteristics of magnetoconcentration photodetector at different temperatures

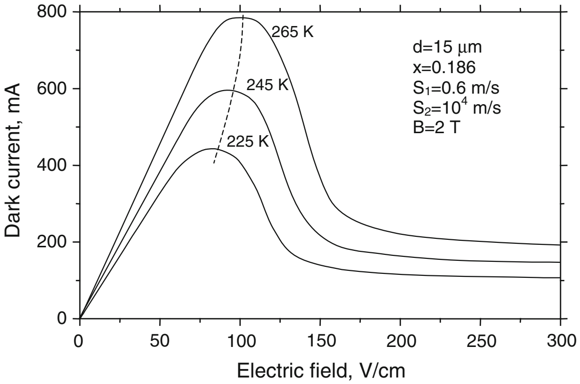


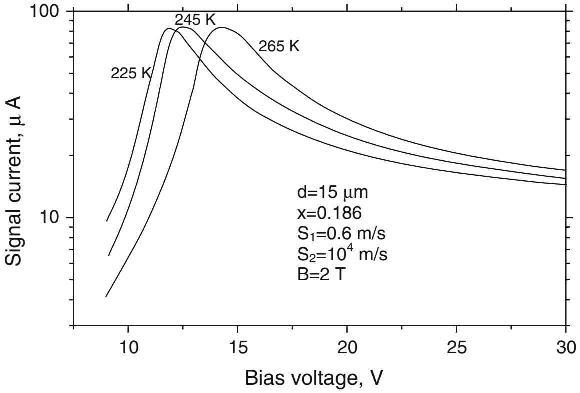
Figure 3.54 shows the calculated UI characteristics of the magnetoconcentration photodetector for $B = 2$ T, $T = 220$ K, $d = 15$ μm, $S_1 = 0.6$ m/s, $S_2 = 10{,}000$ m/s, $x = 0.186$, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $\lambda = 10.6$ μm. The dependence is markedly nonlinear. Only for very low-electric field the dependence is linear (Ohmic). With a field increase, the dependence rises more slowly (sub-Ohmic region) and reaches the maximum where the differential resistance is infinite. For even larger electric fields the region of negative resistance is reached, which is characteristic for all detectors with sufficiently high suppression of Auger processes. At very high field intensities the current becomes saturated. This part of the characteristics belongs to the hot carrier limit, i.e., to the range where the model is not valid anymore.

Magnetic induction follows a similar dependence, first it is linear, then it goes into saturation and at sufficiently high-magnetic fields a negative resistance region appears.

It should be mentioned that any increase of the dynamic resistance and its deviation from the Ohmic dependence means a certain degree of suppression of Auger processes, the larger the nearer is the region of negative resistance.

Figure 3.55 shows the dependence of signal current on bias voltage.

Fig. 3.55 Signal current of a HgCdTe magnetoconcentration detector versus bias



3.7.8 *Noise in Nonequilibrium Magnetoconcentration Photodetectors*

Figure 3.56 shows the photoelectric gain across the thickness of the $Hg_{1-x}Cd_xTe$ magnetoconcentration photodetector for various electric field values. The gain is practically independent on carrier concentration, except near the surface with high recombination rate, where it quickly drops. This is a favorable property of this type of detectors, since it decreases the influence of generation-recombination in the region where this rate reaches its maximum. It is interesting that photoelectric gain decreases with bias.

The calculated photoelectric gain of a magnetoconcentration photoconductor increases with the material composition approaching the optimum at a given temperature and tends to unity, except near the surface with high recombination rate.

Spatial distribution of g-r noise current through a magnetoconcentration detector is shown in Fig. 3.57 for an operating temperature of $T = 220$ K and a magnetic induction of 2 T.

Noise current slowly decreases with the position within the photoconductor, reaches a minimum and then sharply increases toward the back surface (with high recombination velocity.) After reaching its maximum, it becomes lower again because of a sharp drop of photoelectric gain, in spite of the fact that the g-r processes are actually largest just in that position.

The dependence of calculated total noise current through a magnetoconcentration photoconductor on applied bias is shown in Fig. 3.58 for various operating temperatures. To see only the influence of transport parameters to the behavior of the detector, the material composition was kept constant at $x = 0.186$ for all cases shown in Fig. 3.58, although this means a deviation of the device characteristics from the optimum value (responsivity and specific detectivity are obtain for different x at each particular temperature.)

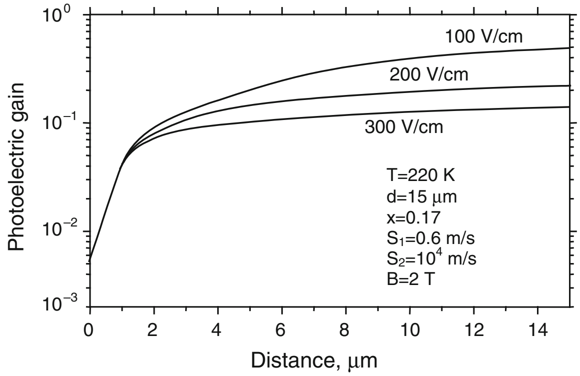


Fig. 3.56 Photoelectric gain across the thickness of a HgCdTe magnetoconcentration photodetector

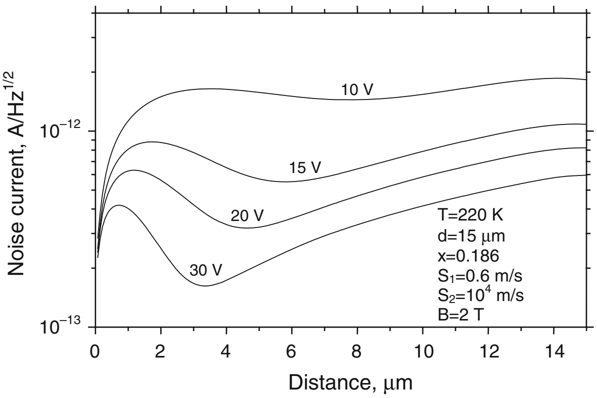
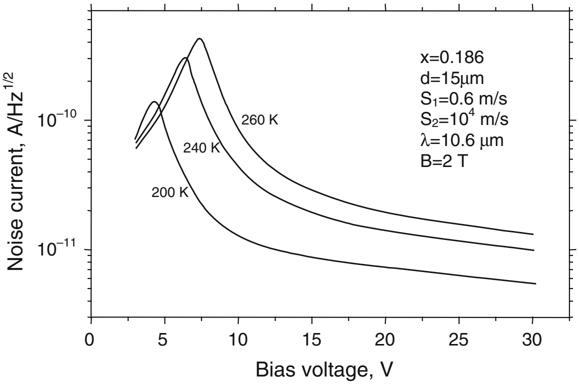


Fig. 3.57 Noise current distribution within a HgCdTe magnetoconcentration photodetector for different values of bias

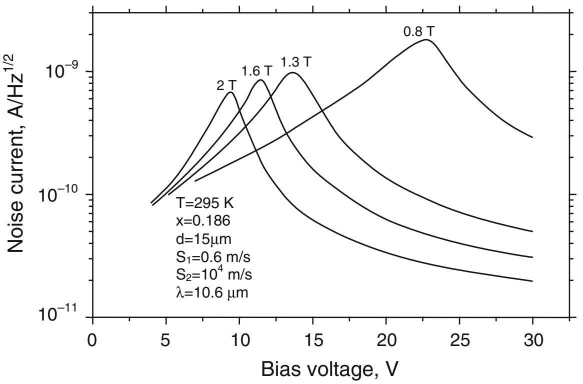
Fig. 3.58 Noise current versus bias for HgCdTe magnetoconcentration photodetector at different temperatures



Noise current reaches a distinct maximum for relatively low bias voltage and subsequently decreases with bias. It should be noted, however, that this is an idealized picture, since in reality a strong additional 1/f component appears, higher with larger voltage.

Figure 3.59 shows the influence of magnetic induction variation to the performance of near-room temperature photodetectors. It can be seen that even relatively small changes of magnetic field strongly change noise current. Noise current significantly increases at lower induction vales (the diagram scale is logarithmic) and reaches higher values in an increasingly wider range of applied voltages. Here too the material composition is not optimized for the operating wavelength, again to allow a more straightforward comparison of detector characteristics to those shown in previous cases.

Fig. 3.59 Noise current versus bias for HgCdTe magnetoconcentration photodetector for different values of magnetic induction



3.7.9 Sensitivity and Detectivity of Nonequilibrium Magnetoconcentration Detectors

Calculated current sensitivity of a magnetoconcentration photoconductor versus bias voltage for different temperatures is shown in Fig. 3.60.

A distinct maximum can be seen at a relatively low bias voltage and temperature only weakly influences its value. A problem is that at similar bias value noise current also reaches its maximum.

Figure 3.61 shows the influence of magnetic field to current sensitivity of the photodetector. An increase of magnetic field dramatically decreases current sensitivity. However, we saw that at the same time it changes noise current, and its total influence works toward improving the detector characteristics.

Specific detectivity versus bias is shown in Fig. 3.62. As expected, it decreases with temperature and improves with bias voltage. It can be also seen that bias voltage influence to specific detectivity is only marked at lower bias values.

Fig. 3.60 Current responsivity of HgCdTe magnetoconcentration photodetector versus bias voltage at different temperatures

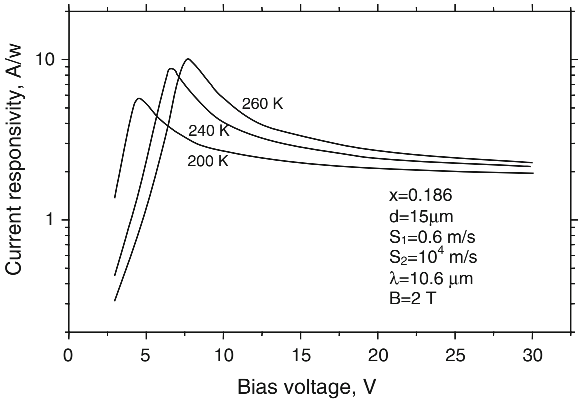


Fig. 3.61 Current responsivity of HgCdTe magnetoconcentration photodetector versus bias voltage for different values of magnetic induction

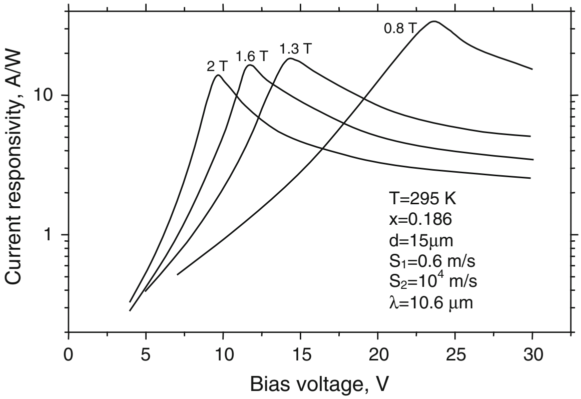


Fig. 3.62 Detectivity of HgCdTe magnetoconcentration photodetector versus bias at different temperatures

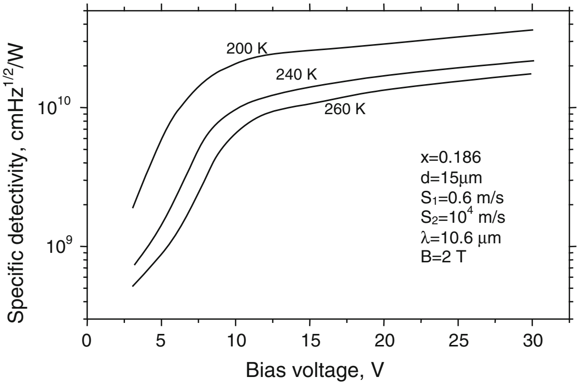


Fig. 3.63 Detectivity of HgCdTe magnetoconcentration photodetector versus bias for different values of magnetic induction

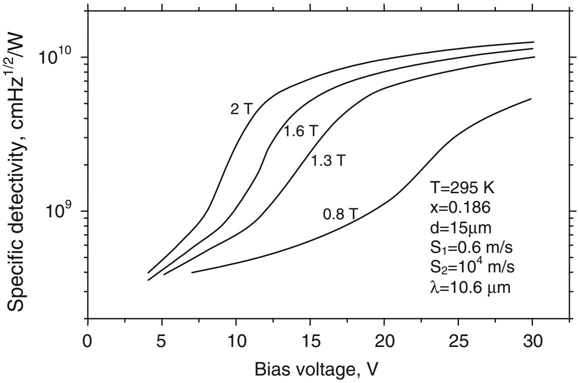


Figure 3.63 shows the influence of magnetic field to specific detectivity of a magnetoconcentration detector. Magnetic induction strongly improves specific detectivity and the required values of bias voltage are lower.

3.7.10 Spectral Characteristics of Nonequilibrium Magnetoconcentration Detectors

The spectral dependence of magnetoconcentration detector sensitivity is determined by spectral absorption coefficient, similar to the conventional photonic detectors. Thus, it follows the well-known triangular shape and increases with wavelength while approaching the absorption edge, and then abruptly drops at the absorption edge. However, the shape is somewhat modified by the inhomogeneous distribution of carriers along the y-axis of the magnetoconcentration detector because of which the slope of the low-wavelength part of the curve is modified. Thus the shape of the curve can be modified to a certain extent by a proper choice of magnetoconcentration parameters (electric field and magnetic induction before all). In this way, a certain degree of control over the parameters of the spectral characteristics is ensured (a dynamic "engineering" or tailoring of the spectral characteristics).

Figure 3.64 shows the spectral current sensitivity of the magnetoconcentration photoconductor at 220 K at a magnetic induction of 2 T and for different electric fields. The maximum of the spectral dependence shifts toward larger wavelength with an increase of the electric field.

Figure 3.65 shows the spectral dependence of specific detectivity of a magnetoconcentration photodetector. Noise is calculated according to the procedure described in Sect. 1.5, without taking into account the 1/f component. Since the g-r noise, being a power function of concentration, decreases with bias faster than current sensitivity, the result is that specific detectivity increases with electric field in spite of the decrease of current sensitivity. Here, the spectral maximum of specific detectivity also shifts with bias, the same as in Fig. 3.64.

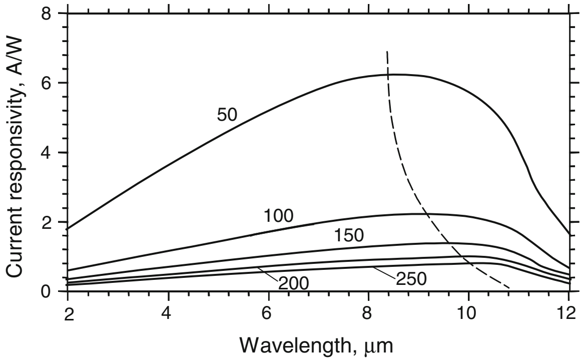


Fig. 3.64 Spectral responsivity of HgCdTe magnetoconcentration photodetector at different electric field values. Numbers near corresponding curves denote electric field in V/cm. $x = 0.186$, $T = 220$ K, $B = 2$ T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 15$ μm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, $\lambda = 10.6$ μm

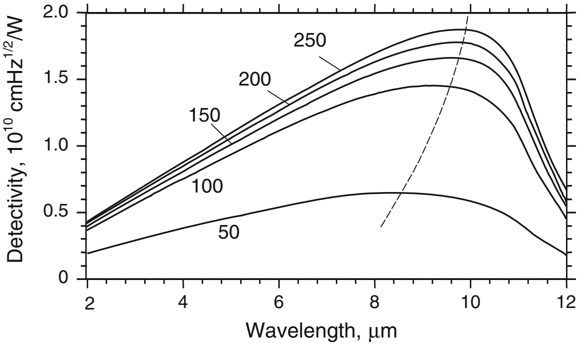


Fig. 3.65 Spectral detectivity of HgCdTe magnetoconcentration photodetector at different electric field values. Numbers near corresponding curves denote electric field in V/cm. Detector parameters are identical to those in Fig. 3.64.

3.7.11 Response Time of Nonequilibrium Magnetoconcentration Photodetector

Approximations usual for the calculation of the frequency response of conventional photonic detectors of optical radiation are not applicable to nonequilibrium detectors, since they operate exclusively under high bias levels. The treatment of magnetoconcentration devices is especially complex in this regard because of the requirement for high magnetic fields. Thus, the approach used here is based on direct numerical solution of the full mathematical-physical model of the device, similarly to the approach to the stationary solution.

To calculate the response time of a magnetoconcentration detectors we utilize the parabolic partial differential equation (3.117) with the coefficients (3.115)–(3.116) and the boundary conditions (3.136)–(3.137), where the expressions (3.121) and (3.122) are used. The concentration profile within the magnetoconcentration photoconductor obtained for a stationary case was used as the initial condition for the determination of the transient solution. The response time was determined as the period between moments of the detector response current reaching 10 % and 90 % of its nominal value (i.e., the stationary value in presence of an optical signal).

Figure 3.66 shows the time evolution of the total current density through a mercury cadmium telluride photoconductor for optical excitation in the form of Heaviside step for the operating temperature 220 K, magnetic induction of 2 T, bias voltage 10 V, illumination arriving from the side depleted by magnetoconcentration. The detector signal current (the difference between the total current of the illuminated detector and the dark current) is 14.4 µA for a detector $l = 1$ mm long, $w = 1$ mm wide, and 10 µm thick at an optical illumination of $P_0 = 10^{-5}$ W. Higher magnetic fields ensure a faster response (the limit being the onset of hot carrier oscillatory effects.)

Figure 3.67 shows time evolution of spatial distribution of current density across the detector. In this case bias is 30 V. Infrared signal arrives from the side with low surface recombination rate.

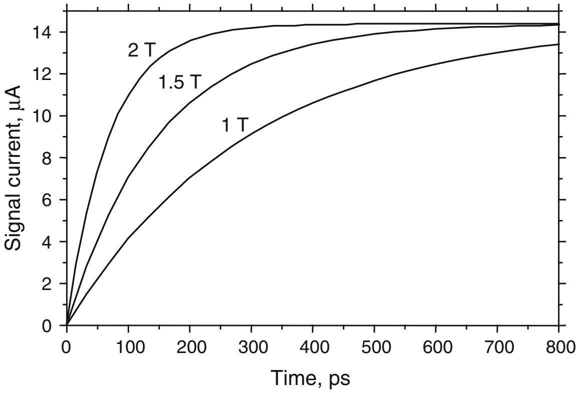


Fig. 3.66 Time evolution of current in HgCdTe magnetoconcentration photodetector for different values of magnetic induction after Heaviside step excitation. x = 0.186, T = 220 K, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, d = 10 μm, S_1 = 1 m/s, $S_2 = 10^4$ m/s, operating wavelength λ = 10.6 μm, bias voltage 10 V, l = 1 mm, w = 1 mm, d = 10 μm, $P_o = 10^{-5}$ W

Figure 3.68 shows the response time versus magnetic induction for a mercury cadmium telluride magnetoconcentration device at two typical operating temperatures (220 K, achievable by thermoelectric coolers and 300 K, room temperature) for $Hg_{1-x}Cd_xTe$ compositions 0.186 and 0.165, respectively (both chosen to enable cutoff wavelength near 10.6 μm for the given temperature).

The response time becomes longer with temperature. The main reason for this is the strong dependence of the response time on carrier mobility. This is a detrimental effect, since the magnetoconcentration detectors are intended for the operation at higher temperatures.

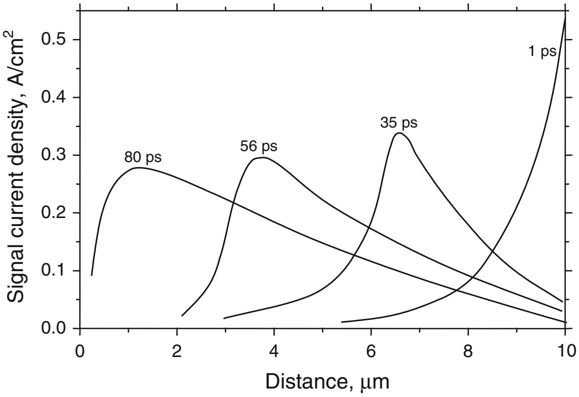


Fig. 3.67 Time evolution of current across the thickness of HgCdTe magnetoconcentration photodetector for different values of magnetic induction after Heaviside step excitation. x = 0.186, T = 220 K, B = 2 T, $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, d = 10 μm, S_1 = 1 m/s, $S_2 = 10^4$ m/s, operating wavelength λ = 10.6 μm, bias voltage 30 V, l = 1 mm, w = 1 mm, d = 10 μm, $P_o = 10^{-5}$ W

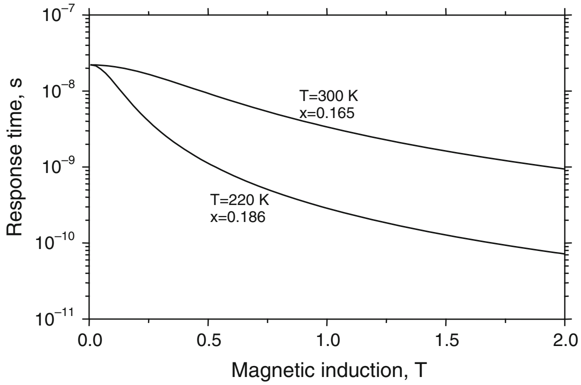


Fig. 3.68 Response time versus magnetic induction in a $Hg_{1-x}Cd_xTe$ magnetoconcentration photodetector at two operating temperatures and for the two corresponding optimum compositions. $N_A = 5 \times 10^{19}$ m^{-3}, $N_D = 1 \times 10^{19}$ m^{-3}, $d = 10$ μm, $S_1 = 1$ m/s, $S_2 = 10^4$ m/s, cutoff wavelength $\lambda = 10.6$ μm, bias voltage 30 V, $l = 1$ mm, $w = 1$ mm, $d = 10$ μm, $P_o = 10^{-5}$ W

On the other hand, the response time becomes shorter with increasing electric and magnetic fields. Theoretically, this means that short response times are achievable at room temperature if the applied fields are sufficiently high. Surely, the necessary condition is to reach some kind of tradeoff between the effect of carrier heating and the shortening of the response time.

For sufficiently high fields the response time becomes shorter than the Auger lifetime and the device effectively operates in the transit time limited mode.

The possibility to tailor the composition of $Hg_{1-x}Cd_xTe$ enables a continuous variation of operating temperatures, applied fields and detector dimensions to reach their optimum values for maximum specific detectivity-bandwidth product. A larger carrier depletion within the active area will not only improve signal-to-noise ratio but simultaneously shorten the response time. However, high fields easily cause transition of carriers to hot region. There is an optimum ratio between the field intensities and the beneficial influence of Auger suppression. This ratio is determined by the characteristic curve drawn between the areas A and B in Fig. 3.46.

3.7.12 Experimental Results on Nonequilibrium Magnetoconcentration Photodetectors

Figure 3.69 shows a measured UI characteristics of an InSb detector at room temperature. As expected, an increase of magnetic induction results in a corresponding increase of the dynamic resistance of the photodetector.

Figure 3.70 shows a ν-type InSb sample reaching negative differential resistance; the dependence is smeared because of the appearance of oscillations.

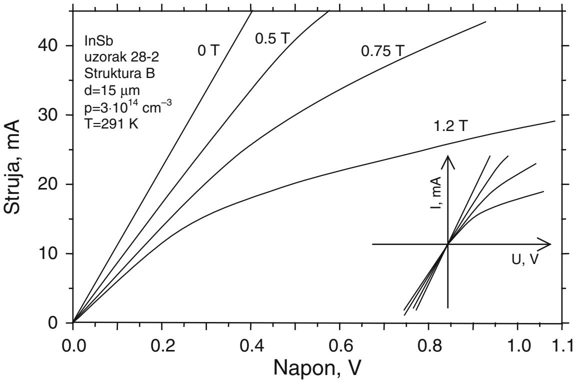
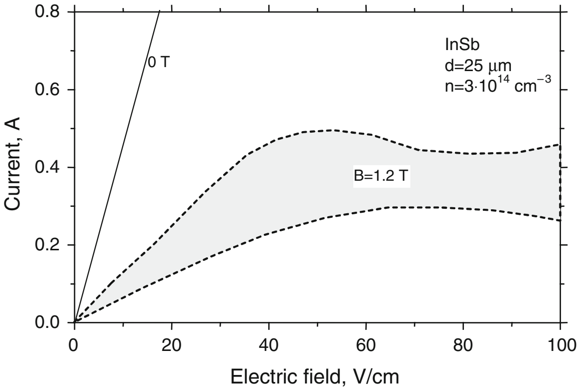


Fig. 3.69 Current-voltage characteristics of a magnetoconcentration InSb detector at 291 K, the direction of the Lorentz force is toward the surface with high recombination rate. The inset shows this dependence for both directions of bias

Fig. 3.70 Experimental current-voltage characteristics of an InSb magnetoconcentration detector



Negative differential resistance and the related high-frequency oscillations in magnetoconcentration InSb structures was theoretically predicted by Guga et al. [396], and experimentally confirmed by Akopyan [395, 402]. These oscillations were observed in photodetectors by Piotrowski, Nowak et al. in mercury cadmium telluride devices [351].

The paper [395] analyses the high-frequency oscillations in InSb magnetoconcentration structure. It is mentioned that the oscillations are complex, with a period below 100 nm and a consideration is given regarding the influence of temperature, doping level, applied fields intensity, and the surface recombination rates at the front and the backside.

Oscillations in the circuits with negative differential resistance are brought into connection with alternating charging and discharging of very small parasitic capacitances, so that the RC constant is also small and thus the oscillation frequency is high. In the course of their work on extraction photodiodes Elliott et al. [361] avoided similar oscillation by connecting in parallel with the detector a resistor with a sufficiently low resistance (7 Ω in their case). The real photodetector current was calculated according to the total measured current.

The nonequilibrium depletion and the increase of dynamic resistance in magnetoconcentration detectors is lower than in the case of heterostructure extraction photodiodes, but significantly better than in exclusion-based isotype devices under similar working conditions (lower dark currents and higher operating temperatures).

Figure 3.71 shows a measured voltage responsivity of an InSb magnetoconcentration photodetector. A significant increase of responsivity is observed when the Lorentz force works in the direction promoting carrier exclusion. According to Malyutenko [397], this is the case if the charge concentration in magnetoconcentration sample is near the intrinsic one. For a reversed direction (accumulation) the sensitivity drops with bias and reaches a level below that of a detector without magnetic field.

An experimental characteristics of spectral noise in an InSb magnetoconcentration detector is shown in Fig. 3.72.

The dependence is typical for semiconductor samples, with the 1/f pink part, a plateau for the g-r noise, and a decreasing high-frequency part which tends to Johnson noise. The knee of the 1/f part is between 10^3 and 10^4 Hz.

It is interesting that the level of the 1/f noise in the device with exclusion direction of the Lorentz force is lower than that in the identical device outside the magnetic field. This means that in the magnetoconcentration devices one may in certain degree tailor the 1/f noise. Similar results were published in [243] for magnetoconcentration samples in lower magnetic fields (below 0.2 T) and at lower temperatures (121 K). This is definitively a comparative advantage of this type of devices over other nonequilibrium types. However, the overall level of this noise component still remains high and is one of the larger problem with magnetoconcentration detector,

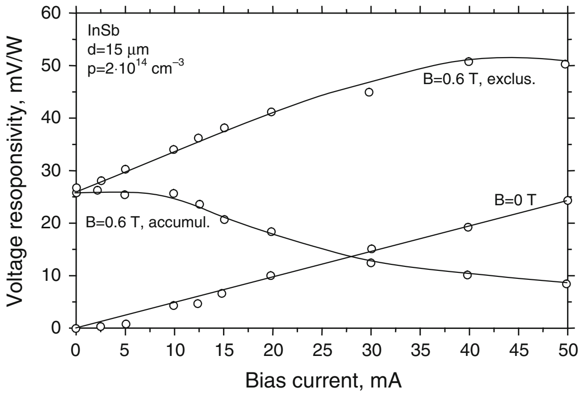


Fig. 3.71 Experimental voltage responsivity of an InSb magnetoconcentration detector for the case of $B = 0$ T and $B = 0.6$ T (direction of Lorentz force toward the surface with higher recombination rate and with lower recombination rate)

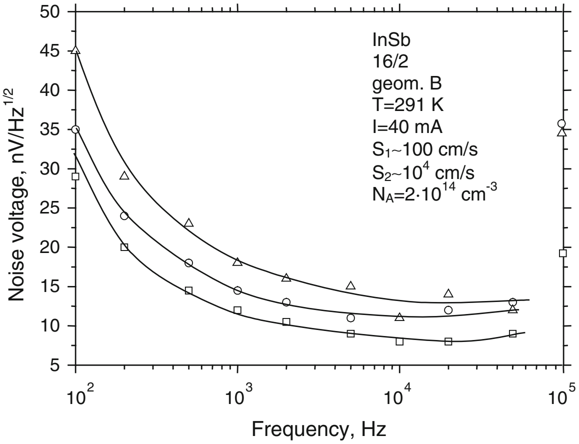


Fig. 3.72 Spectral noise of InSb magnetoconcentration detector at room temperature

similar to other types of devices with nonequilibrium suppression of Auger processes.

3.7.13 Advantages and Shortcomings of Nonequilibrium Magnetoconcentration Photodetectors

According to [403], one of the advantages of the application of magnetic field in photodetectors is that it enables tailoring of spectral shape of sensitivity and detectivity, since magnetic induction determines how much its maximum is going to be enhanced or suppressed. In this way, it is possible to adjust the selectivity of magnetoconcentration devices within certain limits.

Magnetoconcentration photodetectors offer improved dynamic performance under very strong IR illuminations and in the case of high levels of absorption near surface. In conventional detectors in such situations high power is generated in a thin layer of crystal near the surface, which could cause breakdown and the destruction of the device. However, magnetoconcentration drives carriers into the bulk of the device and thus protects it against damages and allows its operation to remain linear. In this way it has improved dynamic range and a built-in protection against accidental “overexposure” by high-power optical signals.

The design of magnetoconcentration photodetectors is very simple, even simpler than an exclusion device. Its basic version is simply a homogeneous slab of narrow-bandgap semiconductor.

The simplicity of the device is an advantage in itself, since besides a simpler technology required for its fabrication it offers much larger space for further refinements. Thus one may expect larger additional performance improvement through application of various mechanisms like e.g., the use of heterojunctions, etc.

An important advantage over other nonequilibrium methods is the possibility to use magnetoconcentration devices to suppress 1/f noise in a certain measure, as described in Sect. 3.7.12.

The most important disadvantages of magnetoconcentration photodetectors are those connected with the very existence of the magnetic field. High magnetic inductions may interfere with the surrounding circuitry even if the fields are well localized, especially if some kind of gyromagnetic instrumentation is present. The confinement of large static magnetic fields in excess of 1 T is a problem in itself.

One of the problems with magnetoconcentration devices is connected with the Joule heating in these devices. As mentioned previously, the inhomogeneous distribution of carriers caused by the Lorentz force concentrates the current flow toward the part with large recombination rate. It is beneficial fact that this surface is directly mounted to a heat sink, thus it is easiest to remove heat generated in its vicinity.

The methods to deal with heat dissipation are well established in the contemporary VLSI technology.

The third group of disadvantages is connected with the requirement to have a surface with an increased recombination rate. This surface introduces additional noise and has a vastly increased density of surface traps. Also, it is necessary to have a relatively accurate and repeatable methods for the fabrication of high- and low surface recombination rates, i.e., to develop technologies for engineering of surface recombination rates.

3.7.14 Basic Implementations of Nonequilibrium Magnetoconcentration Photodetectors

The basic implementation of a nonequilibrium magnetoconcentration photodetector is a plate of homogeneous narrow-bandgap semiconductor, sufficiently thin to avoid oscillations due to cyclotron resonance. Ohmic contacts are formed at the opposite ends of this thin plate. Their purpose is both to bias the detector and to read out the signal (Fig. 3.73a). The bottom surface is chemomechanically polished to achieve a high surface recombination rate, while the recombination rate at the top surface is low. The whole structure is mounted to a metal heat sink.

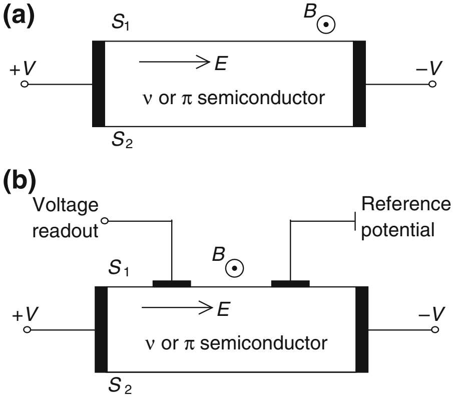


Fig. 3.73 Various implementations of magnetoconcentration photodetector. **a** Standard version; **b** version with separate readout contacts

The structure used in the investigations of Malyutenko et al. (e.g., [397]) is also applicable for the photodetectors with nonequilibrium suppression of Auger processes. This structure is shown in Fig. 3.73b). Here, the voltage readout contacts are located on the surface with low recombination rate and are physically separated from the bias contact pads (which are placed at the ends of the detector, like in the basic version). This structure is more convenient than the basic one because measurements are performed in the part with largest depletion. At the same time, this area absorbs the most part of the incident infrared radiation. All other characteristics are identical to those in the standard version.

The structure shown in Fig. 3.73b is especially convenient for the application with $Hg_{1-x}Cd_xTe$ epitaxial layers whose thickness is of the order of micrometers and whose composition x (and thus their bandgap) varies with the position within the sample, because the compositional nonuniformity can be used for additional decrease of the carrier concentration in the active part.

3.8 Hybrid Methods

Tables 1.6 and 1.7 show that there are families of photonic detectors whose performance could be enhanced through nonequilibrium methods. By a convenient design which would simultaneously utilize magnetoconcentration and some of the other methods, one could expect larger Auger suppression than by applying these methods separately. Electric, magnetic, thermal field inhomogeneities could be used to that purpose.

As an illustration of this approach we consider the case when only electric and magnetic fields are used to enhance the simplest photoconductive and photovoltaic structures. Even in this case two generic combinations appear obvious, together with their subversions: an exclusion photoconductor and extraction-exclusion photodiode, placed into magnetic field so that the Lorentz force promotes the depletion of the low-doped active area.

3.8.1 Magnetoconcentration-Exclusion Device

The simplest hybrid structure is a photodetector with an exclusion contact of $n^+\nu$, $p^+\pi$, $N\nu$, or $P\pi$ type placed in a crossed electric and magnetic field. Figure 3.74 shows as an example an $n^+\nu$ type device. The sample is a thin plate of a narrow-bandgap semiconductor, its thickness comparable to the carrier diffusion length.

The top and the bottom surfaces are chemomechanically treated to ensure low and high surface recombination rate, respectively. Donors are implanted in a thin layer near the positive contact. The bottom surface is mounted to a heat sink. Bias contacts are placed at the beginning of the n^+ and at the end of the ν region, and an additional pair of contacts is placed on the top surface of the ν region (both within the excluded zone) for the readout of the photosignal.

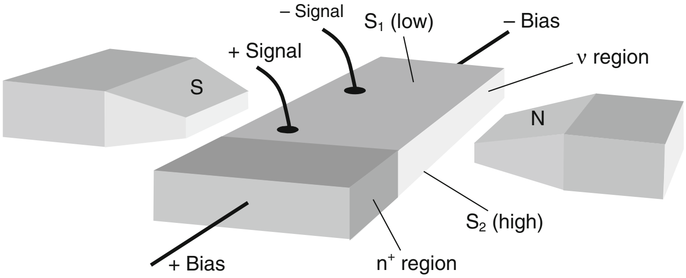


Fig. 3.74 Schematic presentation of a hybrid magnetoconcentration-exclusion photoconductor

The electric field creates the excluded region within the ν-semiconductor near the isotype junction, in a manner analogous to the conventional exclusion device. However, here the Lorentz force additionally drives carriers within the active region toward the high recombination rate backside. It does not influence significantly the carrier distribution within the n^+ (N), part, but strongly modifies the already partially depleted ν-region. In this manner magnetoconcentration complements exclusion and creates lower carrier levels than those achievable through any of the separate mechanisms, without simultaneously causing any additional carrier heating.

3.8.2 Magnetoconcentration-Extraction Device

The concept of the magnetoconcentration-extraction combination (or, more accurately, magnetoconcentration-extraction-exclusion) is actually similar to the one described in the above section. Basically, one places a nonequilibrium extraction diode into a crossed electric and magnetic field, as shown in Fig. 3.75. This device could potentially offer the best performance of all nonequilibrium detectors, since it further enhances the operation of the extraction photodiode which already furnished superior performance.

The geometry of the starting sample is similar to the exclusion case, however, the doping/composition profile is different and corresponds to the typical

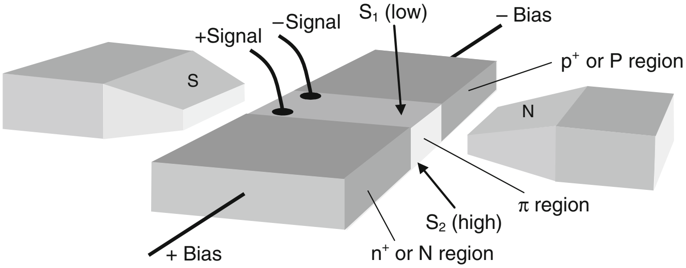


Fig. 3.75 Schematic presentation of a hybrid magnetoconcentration-extraction photodetector

nonequilibrium extraction diode. In this structure the largest magnetoconcentration is reached within the middle, low-doped region. The Lorentz force depletes the part of the active area near the low recombination rate surface. The structure is of "horizontal" type and can be fabricated using the conventional planar technology. The fabrication of alternative structures where the active area is situated between two heterojunctions or the application of even more complex geometries requires accurately controlled epitaxial growth.

3.8.3 Dember Magnetoconcentration Detector

The use of Dember effect for photodetectors at elevated operating temperatures was described in [404]. The same team described a device with simultaneous use of Dember effect and magnetoconcentration [405]. The effect was also described as the giant Dember effect because of high photovoltages obtained in structures utilizing it. A Dember magnetoconcentration photodetector is shown in Fig. 3.76.

Dember effect is essentially the appearance of diffusion photovoltage. In the case of a magnetoconcentration detector, this voltage appears in the direction of the Lorentz force, i.e., perpendicularly to the direction of the vector product of bias and magnetic field. This means that it can be calculated as the difference between the transversal electric field across the illuminated magnetoconcentration device and the same structure without illumination.

The transversal electric field within a magnetoconcentration structure is [405]

$$\begin{aligned} E_y = \frac{1}{q(\mu_p+\mu_n)} \Bigg\{ & \frac{q\mu_n\mu_p(\mu_n+\mu_p)E_xB}{n\mu_{eb}+p\mu_{hb}} - \\ & - \frac{\mu_n\mu_p k_bT}{n\mu_{eb}+p\mu_{hb}}\left[n+p\left(\frac{1}{2}+\frac{1}{4}\frac{p_{ni}}{\sqrt{p_{nb}}}-\frac{\tau_{no}}{2\tau_{po}}\right)\right]\frac{dp}{dy}\left[\frac{\mu_{hb}}{q\mu_n n}-\frac{\mu_{eb}}{q\mu_p p}\right] \\ & + \frac{k_bT}{q}\frac{dp}{dy}\left[\frac{\mu_n}{p}-\frac{\mu_p}{n}\left(\frac{\tau_{n0}}{\tau_{p0}}+\frac{1}{2}\frac{p_{in}}{\sqrt{p_{nb}}}\right)\right]\Bigg\} \end{aligned} \tag{3.154}$$

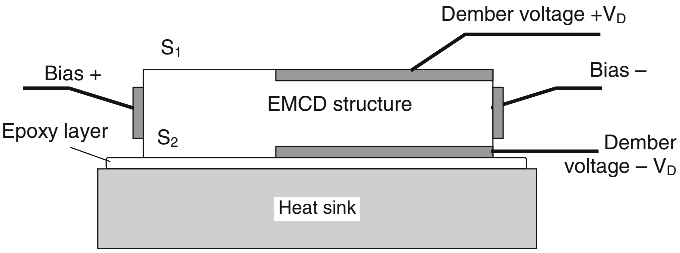
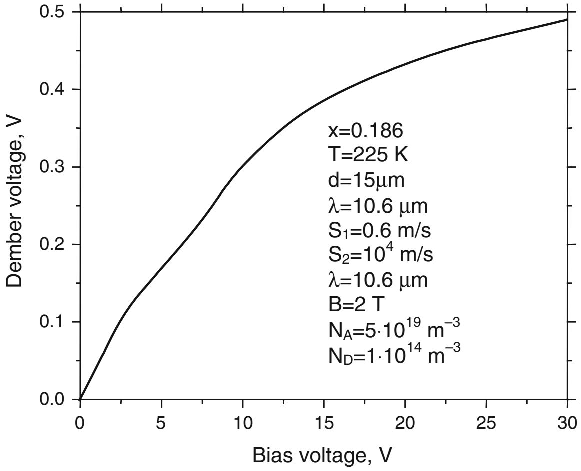


Fig. 3.76 Schematic presentation of Dember-type magnetoconcentration infrared detector

Fig. 3.77 Dependence of Dember photovoltage on bias for mercury cadmium telluride photodetector

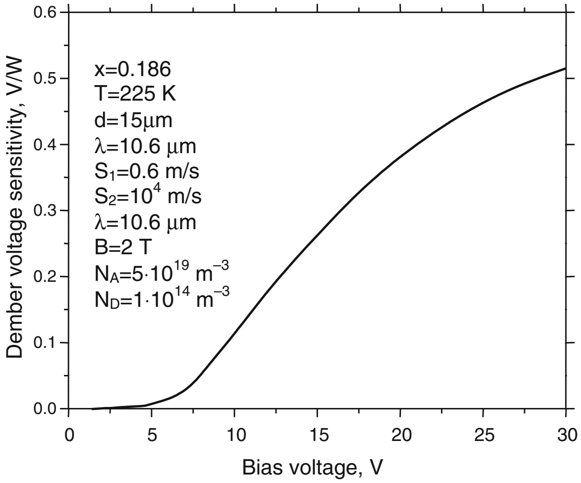


where μ_{eb} and μ_{eb} were previously defined by (3.121) and (3.122), and p_{ni}, p_{nb} are given as (3.124) and (3.125). To determine the integral of this electric field, it is necessary first to calculate the carrier distribution in the structure. This is done by solving the model presented in 3.7.3 in the manner identical to that for the conventional magnetoconcentration (EMCD) detector.

The Dember voltage responsivity is further determined as the ratio between thus obtained voltage and the incident optical signal power. Noise is calculated in the manner identical to the one for the conventional magnetoconcentration detector. Thus, the specific detectivity is fully defined.

Further we consider numerically calculated characteristics of a mercury cadmium telluride Dember magnetoconcentration detector. Figure 3.77 shows the dependence of the Dember voltage on bias, and Fig. 3.78 shows the Dember voltage responsivity versus bias.

Fig. 3.78 Dember voltage responsivity versus bias for HgCdTe magnetoconcentration photodetector



3.9 On Application of Nonequilibrium Detectors: LWIR Free-Space Optics

When nonequilibrium photodetectors first appeared, it seemed apparent that they could revolutionarize the whole field of MWIR and LWIR photodetection and eliminate the needs for cryogenic cooling. However, all of these devices suffer from an excess level of flicker (1/f) noise, with a very high knee frequency, even reaching hundreds of megahertz. This trait reduces or eliminates their applicability in many fields. Yet there is one vast area of application of nonequilibrium IR detectors where low frequency noise is far less important: free-space optics communications in the long-wavelength infrared range.

There are two broad groups of possible last-mile solutions for wireless systems used in building-to-building connections for high-speed metropolitan- or campus-area networks: those utilizing microwaves/millimeter waves and those using infrared (IR) transceivers. The solution of choice for short-range indoor applications are IR wireless systems, and they are well established and standardized (e.g., [406]). The radio systems are better suited for the outdoor use and are thus more often met, but IR systems are also utilized for the same purpose and are offering potentials for much wider bandwidths.

Contemporary infrared (IR) systems for the free-space optics (FSO) in the wireless last-mile access, private networks and mobile infrastructure applications mostly utilize the transmission window 0.7–1.55 μm which is very sensitive to some influences of atmosphere. Rain, fog, smoke, dust particles effectively shorten the available range in this window. Typical fog drops have dimensions comparable to the operating wavelengths, and the resulting real-life maximum distances are typically below 200-300 m, although distances of a few kilometers have been achieved in favorable conditions in point-to-point systems [407].

A possible method to increase the range is to utilize FSO in the atmospheric windows usually connected with military applications such as night vision, surveillance, and targeting and used for decades in bad weather conditions. Probably the most convenient wavelength range is LWIR (Long-Wavelength Infrared Radiation) i.e., (8–14) μm. It is believed that the FSO systems operating in that atmospheric window should be less sensitive to atmospheric disturbances compared to those working at shorter wavelengths. The reason is that longer-lasting fogs usually have smaller drops and thus do not represent an obstacle for longer-wavelength radiation. The technology connected with these systems was described as All-weather Optical Wireless (AOW) communications technology.

The applicability of the night vision windows for last-mile communication was disputed by some authors [408], however, theoretical and experimental observations confirmed that LWIR range indeed offers significant advantages in poor weather conditions. Extensive modeling results for varied real weather conditions based on hourly visibility observations at various airports in the world was presented in [409]. Achour [410] compared FSO at 1.5, 3, 5, and 10 μm. The paper analyzed backscattering, scintillation effects, and overall performance through fog and rain. The conclusion is that performance in fog at 10 μm is significantly better in comparison to shorter wavelength ranges. Colvero et al. performed real-time measurements of free-space optical links to compare their performance at 0.8, 1.6, and 10 μm [243] and came at the identical conclusion. Corrigan et al. performed measurements in 2009 at a multiwavelength test-bed located in the New York metropolitan area [411]. They showed that compared to 1.3 and 1.5 μm the transmission of a link at 8.1 μm in fog is 2–3 times larger and up to 10 times larger after rain and fog event. They attributed this improvement to lower Mie scattering. Hutchinson and Richards [412] also came to a conclusion that FSO at 110 μm perform vastly better than at lower wavelengths. All of these observations are in line with decades of army experience with night vision, surveillance, and targeting systems in bad weather conditions but also by recent measurements dedicated to telecommunication applications.

An important issue regarding use of the night vision infrared wavelengths for the FSO is laser eye safety, as regulated in the IEC Document 60825-1 Amendment 2, since the LWIR FSO are to be used in densely populated areas. The night vision spectrum allows at least 50 times higher laser powers with an even diminished danger to the eye [413] as compared to the 1.55 μm wavelength. Last but not least, the use of night vision wavelengths for public communications is a manner for army applications to find a way for nonmilitary use oriented for the general civil benefit and this alone may be worth the effort.

The other important issue is that this range offers low start-up and operation costs, since this range is not regulated by the Federal Communications Commission (FCC), and does not require expensive licensed spectrum, i.e., it may be utilized anywhere without restrictions [413].

The use of nonequilibrium detectors for all-weather FSO was proposed in [414] and further elaborated in [415].

Conclusion

Microsystem and nanosystem technologies applied to infrared photonic photodetectors open a pathway to a plethora of novel approaches and possibilities for improvement. Some of the concepts that emerged in the past decade enabled methods and techniques that were unimaginable not so long ago. It is sufficient to mention materials with photonic bandgap, metamaterials, plasmonics, but also nonimaging optics and transformation optics to see that the field is very much alive and that novel paradigms do appear in what very recently had appeared a mature and well established area with very little room for further development.

The various methods of photon management through micro and nanophotonics described in Chap. 2 represent a significant part of the answer to the question of the improvement of quantum efficiency and responsivity of semiconductor infrared detectors. They can also help alleviate the problems connected with the influence of background radiation. The methods presented herein represent a mixture of the well-known and well-established methods that were used practically since the inception of infrared detection and the newly introduced methods based on the concepts that appeared very recently. Some of the solutions brought upon by the latter represent a novel dimension that ensures a completely new quality in detector enhancement. To say that they enable thinner and more compact detectors with improved characteristics represents an understatement. The dimensions of the devices can be now reduced much below the operating wavelength limit, and in spite of the deep subwavelength mode of operation not only retain but significantly improve the device performance.

The photon management methods also represent a pathway toward a decrease of generation-recombination processes and the related noise phenomena and pose a viable way to overcome obstacles that until recently appeared insurmountable. The mechanism of photon recycling by cavity enhancement ensures direct suppression of radiative noise, while the possibility to localize fields at a subwavelength level ensures vastly smaller photodetector volumes, thus helping overcome problems with Auger and Shockley-Read phenomena as well.

The nonequilibrium methods are the other part of the reply to the question of noise suppression. They proved themselves a means to suppress Auger recombination, but also for a significant part the Auger generation as well. Each

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2

of the three generic groups of structures for nonequilibrium suppression has its advantages and disadvantages: photoconductors with isotype junctions in external electric field (exclusion devices), reversely biased p-n junctions combined with isotype junctions at a distance below the diffusion length (extraction diodes) and photodetectors subject to Lorentz force, i.e., placed in crossed electric and magnetic field (magnetoconcentration devices).

A possible classification of the methods for infrared detector enhancement is proposed in this book. Besides relating the existing strategies for detector improvement, both the old and the more recent ones, it also points out to possible solutions that could be implemented in the future. The largest further improvements could be expected from the photon management methods, and especially from those based on recently proposed concepts. Especially, the development of nanotechnologies and nanophotonics proved fruitful to that purpose. One could mention the PCE (photonic crystal enhancement), nonimaging optics, transformation optics, metamaterials, and plasmonics. Not much more than a decade ago researchers debated if it is possible at all to have values of refractive index below zero and if one can fabricate a superlens overcoming the diffraction limit. Now plasmonics and metamaterials ensure field confinements in subwavelength volumes, operation below the Abe limit, and offer a possibility of tailoring of dispersion relation almost at will. It is reasonable to assume that future research in this living field will bring further advancements. Obviously, the possibilities of use of many of the mentioned methods and technologies yet remains to be properly investigated, since many of the concepts did not exist at all until very recently.

Microsystem technologies (microfabrication or MEMS technologies) and the subsequently developed nanofabrication techniques appeared basically as a extension and upgrade of planar technologies which were used for the fabrication of photonic detectors since their very beginnings. It is actually logical that these new technologies in their turn are ensuring the fabrication of the structures and systems to improve and enhance these photonic devices.

Since the structures and processes used for detector enhancement are interconnected, it often occurs that an approach introduced because of one effect also brings to an improvement of another one. A typical example can be found with methods for optical trapping that were introduced to improve the optical path through the detector (absorption coefficient-thickness product), to turn out that these methods not only improve radiative lifetime through photon recycling but even Auger and Shockley-Read lifetimes due to a decrease of overall dimensions of the detector.

A large impetus to the field of photodetector enhancement came from the researchers working on the improvement of solar cells. These days a large scientific input related with these energy sources is met within the context of plasmonics. There is obviously a synergy between different research fields related with photodetection and generally with nanophotonics and solutions that appeared in one of the subfield are applicable in other related fields as well.

An important conclusion to be drawn from the present text is that a combination of equilibrium and nonequilibrium methods has potentials not only to reach the BLIP limit, but also to surpass it, nearing the contemporary infrared devices to the signal fluctuation limited mode of operation. There are no fundamental limits that would hinder novel generations of infrared photonic detectors to overcome the need for cooling. Of all types of photodetectors quoted in Part I, the photonic devices remain the closest to the imaginary ideal device for detection of infrared radiation.

Finally, the majority of the photon management methods presented herein can be applied with little or no modification to other types of detectors. Among these, probably the most important devices nowadays are solar cells.

References

1. R. Hudson, *Infrared System Engineering* (Wiley, Hoboken, 1969)
2. R.C. Jones, Phenomenological Description of the Response and Detecting Ability of Radiation Detectors. Proc. IRE **47**(9), 1495–1502 (1959)
3. R.J. Keyes (ed.), *Optical and Infrared Detectors* (Springer Verlag, Berlin, 1983)
4. P.N.J. Dennis, *Photodetectors: An Introduction to Current Technology* (Plenum Press, New York, 1987)
5. P. Kruse, L. McGlauchlin, R. McQuistan, *Elements of Infrared Technology: Generation, Transmission and Detection* (Wiley, New York, 1962)
6. E.H. Putley, Thermal Detectors, in *Optical and Infrared Detectors,* ed. by R.J. Keyes (Springer, Berlin, 1983)
7. D. Long, Photovoltaic and Photoconductive Infrared Detectors, in *Optical Infrared Detectors*, ed by R.J. Keyes (Springer, Berlin, 1983)
8. A. Rogalski, *Infrared Detectors* (CRC Press, Bocca Raton, 2010)
9. J. Piotrowski, A. Rogalski, *High-Operating-Temperature Infrared Photodetectors* (SPIE Press, Bellingham, 2007)
10. Z. Djurić, Quantum efficiency of photoconductive detectors-influence of reflection and surface recombination velocity. Infrared Phys. **27**(6), 407–410 (1987)
11. Z. Djurić, Z. Jakšić, Back side reflection influence on quantum efficiency of photovoltaic devices. Electronics Lett. **24**(17), 1100–1101 (1988)
12. V. Gopal, Surface recombination in photoconductors. Infrared Phys. **25**(4), 615–618 (1985)
13. A. Rogalski, J. Piotrowski, Intrinsic infrared detectors. Progress Quantum Electron. **12**(2–3), 87–289 (1988)
14. I.M. Baker, F.A. Capocci, D.E. Charlton, J.T.M. Wotherspoon, Recombination in cadmium mercury telluride photodetectors. Solid State Electron. **21**(11–12), 1475–1480 (1978)
15. B.L. Gel'mont, K.S. Kim, M. Shur, Theory of impact ionization and auger recombination in $Hg_{1-x}Cd_xTe$. Phys. Rev. Lett. **69**(8), 1280–1282 (1992)
16. B.L. Gel'mont, Auger Recombination in Narrow-Gap Semiconductors. Sov. Phys. Semiconductors **14**(10), 1140–1142 (1980)
17. W. Van Roosbroeck, W. Shockley, Photon-radiative recombination of electrons and holes in germanium. Phys. Rev. **94**(6), 1558–1560 (1954)
18. E. Finkman, S.E. Schacham, The exponential optical absorption band tail of $Hg_{1-x}Cd_xTe$. J. Appl. Phys. **56**(10), 2896–2900 (1984)
19. M.A. Kinch, M.J. Brau, A. Simmons, Recombination mechanisms in 8-14-μ HgCdTe. J. Appl. Phys. **44**(4), 1649–1663 (1973)
20. A. Rogalski, Z. Orman, Band-to-band recombination in $InAs_{1-x}Sb_x$. Infrared Phys. **25**(3), 551–560 (1985)
21. D.L. Polla, R.L. Aggarwal, J.A. Mroczkowski, J.F. Shanley, M.B. Reine, Observation of deep levels in $Hg_{1-x}Cd_xTe$ with optical modulation spectroscopy. Appl. Phys. Lett. **40**(4), 338–340 (1982)

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2

22. R.G. Pratt, J. Hewett, P. Capper, C.L. Jones, M.J. Quelch, Minority carrier lifetime in n-type Bridgman grown $Hg_{1-x}Cd_xTe$. J. Appl. Phys. **54**(9), 5152–5157 (1983)
23. Z. Djurić, Z. Jakšić, Z. Djinović, M. Matić, Ž. Lazić, Some theoretical and technological aspects of uncooled HgCdTe detectors: a review. Microelectron. J. **25**(2), 99–114 (1994)
24. Z. Jakšić, O. Jakšić, Simple approximation for absorption coefficient in degenerate HgCdTe, pp. 117–120
25. G.L. Hansen, J.L. Schmit, T.N. Casselman, Energy gap versus alloy composition and temperature in $Hg_{1-x}Cd_xTe$. J. Appl. Phys. **53**(10), 7099–7101 (1982)
26. Z. Jakšić, O. Jakšić, Dispersion of refractive index in degenerate mercury cadmium telluride, pp. 95–98
27. W.P. Dumke, Spontaneous radiative recombination in semiconductors. Phys. Rev. **105**(1), 139–144 (1957)
28. R.G. Humphreys, Radiative lifetime in semiconductors for infrared detection. Infrared Phys. **23**(3), 171–175 (1983)
29. R.G. Humphreys, Radiative lifetime in semiconductors for infrared detection. Infrared Phys. **26**(6), 337–342 (1986)
30. P.T. Landsberg, D.J. Robbins, The first 70 semiconductor Auger processes. Solid State Electron. **21**(11–12), 1289–1294 (1978)
31. A.R. Beattie, P.T. Landsberg, Auger effect in semiconductors. Proc. Roy. Soc. Lond. Ser. A. Math. Phys. Sci. **249**(1256), 16–29 (1959)
32. A.R. Beattie, G. Smith, Recombination in semiconductors by a light hole auger transition, Phys. Status Solidi B **19**(2), 577–586 (1967)
33. M. Takeshima, Auger recombination in InAs, GaSb, InP, and GaAs. J. Appl. Phys. **43**(10), 4114–4119 (1972)
34. M. Takeshima, Effect of anisotropic band parameters on band-to-band Auger recombination in $In_{0.72}Ga_{0.28}As_{0.6}P_{0.4}$. Phys. Rev. B **29**(4), 1993–2001 (1984)
35. A. Sugimura, Band-to-band Auger effect in GaSb and InAs lasers. J. Appl. Phys. **51**(8), 4405–4411 (1980)
36. A. Sugimura, Band-to-band Auger effect in long wavelength multinary III-V alloy semiconductor lasers. IEEE J. Quantum Electron. **18**(3), 352–363 (1982)
37. B.L. Gel'mont, Z.N. Sokolova, I.N. Yassievich, Auger Recombination in Direct-Gap p-Type Semiconductors. Sov. Phys. Semiconductors **16**(4), 382–387 (1982)
38. A. Haug, Auger recombination in direct-gap semiconductors: band-structure effects. J. Phys. C: Solid State Phys. **16**(21), 4159–4172 (1983)
39. A. Haug, Band-to-band Auger recombination in semiconductors. J.Phys. Chem. Solids **49**(6), 599–605 (1988)
40. R.R. Gerhardts, R. Dornhaus, G. Nimtz, The Auger-effect in $Hg_{1-x}Cd_xTe$. Solid State Electron. **21**(11–12), 1467–1470 (1978)
41. T.N. Casselman, Calculation of the auger lifetime in p-type $Hg_{1-x}Cd_xTe$. J. Appl. Phys. **52** (2), 848–854 (1981)
42. F. Bertazzi, M. Goano, E. Bellotti, Calculation of auger lifetimes in HgCdTe. J. Electron. Mat. **40**(8), 1663–1667 (2011)
43. P.T. Landsberg, *Recombination in Semiconductors* (Cambridge University Press, Cambridge, 2003)
44. J.S. Blakemore, *Semiconductor Statistics* (Dover Publications, Mineola, New York, 2002)
45. A.R. Beattie, P.T. Landsberg, One-dimensional overlap functions and their application to auger recombination in semiconductors. Proc. Royal Soc. London. Ser. A. Math. Phys. Sci. **258**(1295), 486–495 (1960)
46. E. Antončik, P.T. Landsberg, Overlap integrals for Bloch electrons. Proc. Phys. Soc. **82**(3), 337–342 (1963)
47. M.G. Burt, S. Brand, C. Smith, R.A. Abram, Overlap integrals for Auger recombination in direct-bandgap semiconductors: calculations for conduction and heavy-hole bands in GaAs and InP. J. Phy. C: Solid State Phys. **17**(35), 6385–6401 (1984)

48. A.R. Beattie, P. Scharoch, R.A. Abram, Impact ionisation threshold energy surfaces for anisotropic band structures in semiconductors. Semicond. Sci. Tech. **4**(9), 715–723 (1989)
49. A.R. Beattle, R.A. Abram, P. Scharoch, Realistic evaluation of impact ionisation and Auger recombination rates for the ccch transition in InSb and InGaAsP. Semicond. Sci. Tech. **5**(7), 738–744 (1990)
50. A.M. White, The characteristics of minority-carrier exclusion in narrow direct gap semiconductors. Infrared Phys. **25**(6), 729–741 (1985)
51. Z. Djurić, Z. Jakšić, A. Vujanić, M. Smiljanić, A simple approximative method for determination of Auger 1 lifetime in degenerate narrow gap semiconductors. Infrared Phys. **34**(6), 601–605 (1993)
52. J. Camassel, D. Auvergne, H. Mathieu, Temperature dependence of the band gap and comparison with the threshold frequency of pure GaAs lasers. J. Appl. Phys. **46**(6), 2683–2689 (1975)
53. D. Hahn, O. Jaschinski, H.H. Wehmann, A. Schlachetzki, M. Von Ortenberg, Electron-concentration dependence of absorption and refraction in n-$In_{0.53}Ga_{0.47}As$ near the band-edge. J. Electron. Mat. **24**(10), 1357–1361 (1995)
54. A.R. Beattie, A.M. White, An analytic approximation with a wide range of applicability for electron initiated Auger transitions in narrow-gap semiconductors. J. Appl. Phys. **79**(2), 802–813 (1996)
55. W. Shockley, W.T. Read, Statistics of the recombinations of holes and electrons. Phys. Rev. **87**(5), 835–842 (1952)
56. R.N. Hall, Electron-hole recombination in germanium. Phys. Rev. **87**(2), 387–387 (1952)
57. Y.L. Tyan, T.R. Schimert, L.T. Claiborne, Analysis of excess carrier lifetime in p-type HgCdTe using a three-level shockley-read model. J. Vac. Sci. Technol. B **10**(4), 1560–1568 (1992)
58. S.M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981)
59. J.B. Johnson, Thermal agitation of electricity in conductors. Phys. Rev. **32**(1), 97–109 (1928)
60. H. Nyquist, Thermal agitation of electric charge in conductors. Phys. Rev. **32**(1), 110–113 (1928)
61. M.S. Gupta, *Electrical Noise: Fundamentals and Sources* (IEEE Press, New York, 1977)
62. P. Bhattacharya, *Semiconductor Optoelectronic Devices* (Prentice Hall, New Jersey, 1997)
63. J. Singh, *Semiconductor Optoelectronics* (McGraw Hill, New York, 1995)
64. A. Ambrozy, *Electronic Noise* (McGraw Hill, New York, 1982)
65. K.M. Van Vliet, Noise in Semiconductors and Photoconductors. Proc. IRE **46**(6), 1004–1018 (1958)
66. R.E. Burgess, Electronic fluctuations in semiconductors. Brit. J. Appl. Phys. **6**(6), 185–190 (1955)
67. Z. Djurić, Z. Jakšić, A. Vujanić, J. Piotrowski, Auger generation suppression in narrow-gap semiconductors using the magnetoconcentration effect. J. Appl. Phys. **71**(11), 5706–5708 (1992)
68. N.B. Luk'yanchikova, N.P. Garbar, V.K. Malyutenko, G.I. Teslenko, Fluctuations of the Conductivity of Indium Antimonide in Crossed Electric and Magnetic Fields. Sov. Phys. Semiconductors **15**(2), 193–198 (1981)
69. J. Piotrowski, K. Adamiec, A. Maciak, High-temperature 10.6μm HgZnTe photodetectors. Infrared Phys. **29**(2–4), 267–270 (1989)
70. Z. Jakšić, Z. Djurić, Z. Djinović, R. Petrović, Composition profiles versus growth pressure and temperature in epitaxial HgZnTe layers. Phys. Status Solidi A. **152**(2), 451–459 (1995)
71. A. Rogalski, $Hg_{1-x}Zn_xTe$ as a potential infrared detector material. Prog. Quantum Electron. **13**(4), 299–353 (1989)
72. W.W. Anderson, Tunnel contribution to $Hg_{1-x}Cd_xTe$ and $Pb_{1-x}Sn_xTe$ p-n junction diode characteristics. Infrared Phys. **20**(6), 353–361 (1980)
73. P.R. Emtage, Auger recombination and junction resistance in lead-tin telluride. J. Appl. Phys. **47**(6), 2565–2568 (1976)

74. A. Rogalski, $InAs_{1-x}Sb_x$ infrared detectors. Prog. Quantum Electron. **13**(3), 191–231 (1989)
75. P.G. Datskos, N.V. Lavrik, S. Rajic, Performance of uncooled microcantilever thermal detectors. Rev. Sci. Instrum. **75**(4), 1134–1148 (2004)
76. T. Trupke, M.A. Green, P. Würfel, Improving solar cell efficiencies by up-conversion of sub-band-gap light. J. Appl. Phys. **92**(7), 4117–4122 (2002)
77. T. Trupke, M.A. Green, P. Würfel, Improving solar cell efficiencies by down-conversion of high-energy photons. J. Appl. Phys. **92**(3), 1668–1674 (2002)
78. S.J. Fonash, *Solar Cell Device Physics* (Elsevier, Amsterdam, 2010)
79. G. Li, R. Zhu, Y. Yang, Polymer solar cells. Nat. Photonics **6**(3), 153–161 (2012)
80. A. McEvoy, T. Markvart, L. Castañer, *Solar Cells* (Elsevier, Amsterdam, 2013)
81. A. Shah, P. Torres, R. Tscharner, N. Wyrsch, H. Keppner, Photovoltaic technology: the case for thin-film solar cells, Science **285**(5428), 692–698 (1999)
82. E. Yablonovitch, Statistical ray optics. J. Opt. Soc. Am. **72**, 899–907 (1982)
83. E. Yablonovitch, G.D. Cody, Intensity enhancement in textured optical sheets for solar cells. IEEE Trans. Electron Devices **ED-29**(2), 300–305 (1982)
84. T. Tiedje, E. Yablonovitch, G.D. Cody, B.G. Brooks, Limiting efficiency of silicon solar cells, IEEE Trans. Electron Devices **ED-31**(5), 711–716 (1984)
85. P. Campbell, M.A. Green, Limiting efficiency of silicon solar cells under concentrated sunlight. IEEE Trans. Electron Devices **ED-33**(2), 234–239 (1986)
86. W. Shockley, H.J. Queisser, Detailed balance limit of efficiency of p-n junction solar cells. J. Appl. Phys. **32**(3), 510–519 (1961)
87. M. Born, E. Wolf, *Principles of Optics*, 7th edn. (Cambridge University Press, Cambridge, 1999)
88. Z. Yu, A. Raman, S. Fan, Fundamental limit of nanophotonic light trapping in solar cells. Proc. Nat. Acad. Sci. U.S.A. **107**(41), 17491–17496 (2010)
89. Z. Yu, A. Raman, S. Fan, Thermodynamic upper bound on broadband light coupling with photonic structures, Phys. Rev. Lett. **109**(17) 173901 (2012)
90. D. M. Callahan, J.N. Munday, H.A. Atwater, Solar cell light trapping beyond the ray optic limit. Nano Lett. **12**(1), 214–218 (2012)
91. V. Ganapati, O.D. Miller, E. Yablonovitch, Light trapping textures designed by electromagnetic optimization for subwavelength thick solar cells. IEEE J.Photovoltaics **4**, 175–182 (2013)
92. J. Chaves, *Introduction to Nonimaging Optics* (CRC Press, Bocca Raton, 2008)
93. J.J. O'Galagher, *Nonimaging optics in Solar Energy* (Morgan and Claypool, San Rafael, 2008)
94. W.T. Welford, R. Winston, *High Collection Nonimaging Optics* (Academic Press, New York, 1989)
95. R. Winston, Principles of solar concentrators of a novel design. Sol. Energy **16**(2), 89–95 (1974)
96. R. Winston, Dielectric compound parabolic concentrators. Appl. Opt. **15**(2), 291–293 (1976)
97. R. Winston, J.C. Minano, P.G. Benitez, *Nonimaging Optics* (Academic Press, New York, 2005)
98. K.S. Lee, F.S. Barnes, Microlenses on the end of single-mode optical fibers for laser applications. Appl. Opt. **24**(19), 3134–3139 (1985)
99. B. Hillerich, Influence of lens imperfections with LD and LED to single-mode fiber coupling. J. Lightwave Technol. **7**(1), 77–86 (1989)
100. K.S. Lee, Focusing characteristics of a truncated and aberrated Gaussian beam through a hemispherical microlens. Appl. Opt. **25**(20), 3671–3676 (1986)
101. L.G. Cohen, M.V. Schneider, Microlenses for Coupling Junction Lasers to Optical Fibers. Appl. Opt. **13**(1), 89–94 (1974)
102. H. Karstensen, R. Frankenberger, High-efficiency two lens laser diode to single-mode fiber coupler with a silicon plano convex lens. J. Lightwave Technol. **7**(2), 244–249 (1989)

103. A. Nicia, Lens coupling in fiber-optic devices: efficiency limits. Appl. Opt. **20**(18), 3136–3145 (1981)
104. V. Russo, G.C. Righini, S. Sottini, S. Trigari, Lens-ended fibers for medical applications: a new fabrication technique. Appl. Opt. **23**(19), 3277–3283 (1984)
105. J. Sakai, T. Kimura, Design of a miniature lens for semiconductor laser to single-mode fiber coupling. IEEE J. Quantum Electron. **16**(10), 1059–1067 (1980)
106. M. Saruwatari, K. Nawata, Semiconductor laser to single-mode fiber coupler. Appl. Opt. **18** (11), 1847–1856 (1979)
107. P. Kayoun, C. Puech, M. Papuchon, H. Arditty, Improved coupling between laser diode and single-mode fibre tipped with a chemically etched self-centred diffracting element. Electron. Lett. **17**(12), 400–402 (1981)
108. H. Sakaguchi, N. Seki, S. Yamamoto, Power coupling from laser diodes into single-mode fibres with quadrangular pyramid-shaped hemiellipsoidal ends. Electron. Lett. **17**(12), 425–426 (1981)
109. S. Mukai, M. Watanabe, H. Itoh, H. Yajima, Integration of a diode laser and an electronic lens for controlling the beam focus position. Appl. Phys. Lett. **54**, 315 (1989)
110. B. Hillerich, Influence of lens imperfection with LD and LED to single-mode fiber coupling. Lightwave Technol. J. **7**(1), 77–86 (1989)
111. S.-D. Moon, N. Lee, S. Kang, Fabrication of a microlens array using micro-compression molding with an electroformed mold insert. J. Micromech. Microeng. **13**(1), 98 (2003)
112. D. Kato, Light coupling from a stripe-geometry GaAs diode laser into an optical fiber with spherical end. J. Appl. Phys. **44**(6), 2756–2758 (1973)
113. Z.D. Popovic, R.A. Sprague, G. Connell, Technique for monolithic fabrication of microlens arrays. Appl. Opt. **27**(7), 1281–1284 (1988)
114. Y. Odagiri, M. Shikada, K. Kobayashi, High-efficiency laser-to-fibre coupling circuit using a combination of a cylindrical lens and a selfoc lens. Electron. Lett. **13**(14), 395–396 (1977)
115. J. Heinen, Preparation and properties of monolithically integrated lenses on InGaAsP/InP light-emitting diodes. Electron. Lett. **18**(19), 831–832 (1982)
116. M. W. Farn, W. B. Veldkamp, Binary optics. Handb. Opt. **2**, 8.1–8.19 (1995)
117. F. Ostermayer, P. Kohl, R. Burton, Photoelectrochemical etching of integral lenses on InGaAsP/InP light-emitting diodes. Appl. Phys. Lett. **43**(7), 642–644 (1983)
118. M.C. Wu, Micromachining for optical and optoelectronic systems. Proc. IEEE **85**(11), 1833–1856 (1997)
119. Y. Ono, Y. Kimura, Y. Ohta, N. Nishida, Antireflection effect in ultrahigh spatial-frequency holographic relief gratings. Appl. Opt. **26**(6), 1142–1146 (1987)
120. H.M. Smith, *Holographic Recording Materials* (Springer-Verlag, Berlin, 1977)
121. L. Huff, M. Bass, E.W.V. Stryland, D.R. Williams, W.L. Wolfe (eds.), in *Handbook of Optics*, Holography and Holographic Instruments (McGraw-Hill, New York, 1995), pp. 23.21–23.31
122. N.L. Kazanskiy, V.V. Kotlyar, V.A. Soifer, Computer-aided design of diffractive optical elements. Opt. Eng. **33**(10), 3156–3166 (1994)
123. S. Calixto, G.P. Padilla, Micromirrors and microlenses fabricated on polymer materials by means of infrared radiation. Appl. Opt. **35**(31), 6126–6130 (1996)
124. A. Yi, L. Li, Design and fabrication of a microlens array by use of a slow tool servo. Opt. Lett. **30**(13), 1707–1709 (2005)
125. G. Milton, Y. Gharbia, J. Katupitiya, Mechanical fabrication of precision microlenses on optical fiber endfaces. Opt. Eng. **44**(12) 123402–123408 (2005)
126. M. Grudzien, J. Piotrowski, Monolithic optically immersed HgCdTe IR detectors. Infrared Phys. **29**(2), 251–253 (1989)
127. I.M. Bassett, W.T. Welford, R. Winston,in *Progress in Optics*, ed by E. Wolf. Nonimaging Optics for Flux Concentration, vol 27 (Elsevier,Amsterdam, 1989), pp. 161–226
128. R. Winston, J.C. Miñano, P. Benítez, *Nonimaging Optics* (Elsevier Academic Press, Burlington, 2005)

129. A. Polman, H.A. Atwater, Photonic design principles for ultrahigh-efficiency photovoltaics. Nat. Mater. **11**(3), 174–177 (2012)
130. J.H. Atwater, P. Spinelli, E. Kosten, J. Parsons, C. Van Lare, J. Van De Groep, J. Garcia De Abajo, A. Polman, H.A. Atwater, Microphotonic parabolic light directors fabricated by two-photon lithography. Appl. Phys. Lett. **99**(15) 151113 (2011)
131. Y.-L. He, J. Xiao, Z.-D. Cheng, Y.-B. Tao, A MCRT and FVM coupled simulation method for energy conversion process in parabolic trough solar collector. Renewable Energy **36**(3), 976–985 (2011)
132. D. Feuermann, J. Gordon, Analysis of a Two-Stage Linear Fresnel Reflector Solar Concentrator. J. Sol. Energy Eng. US **113**(4) 272–279 1991
133. B. Negi, S. Mathur, T. Kandpal, Optical and thermal performance evaluation of a linear Fresnel reflector solar concentrator. Sol. Wind Technol. **6**(5), 589–593 (1989)
134. O.K. Ersoy, *Diffraction, Fourier Optics and Imaging* (Wiley, Hoboken, 2007)
135. E. Hecht, A. Zajac, *Optics* (Addison-Wesley, Boston, MA, USA, 1990)
136. C. Wu Ming, Micromachining for optical and optoelectronic systems, Proc. IEEE **85**(11), 1833–1856 (1997)
137. H.J. Caulfield (ed.), *Handbook of Optical Holography* (Academic Press, Waltham, 1979)
138. D.C. O'Shea, T.J. Suleski, A.D. Kathman, D.W. Prather, *Diffractive Optics: Design, Fabrication, and Test* (SPIE Publications, Bellingham, 2003)
139. J.A. Jordan Jr, P.M. Hirsch, L.B. Lesem, D.L. Van Jooy, Kinoform Lenses. Appl. Opt. **9**(8), 1883–1887 (1970)
140. J.R. Wendt, G.A. Vawter, R.E. Smith, M.E. Warren, Fabrication of subwavelength, binary, antireflection surface-relief structures in the near infrared. J. Vac. Sci. Technol. B: Microelectron. Nanometer Struct. **14**(6), 4096–4099 (1996)
141. G. Tricoles, Computer generated holograms: an historical review. Appl. Opt. **26**(20), 4351–4357 (1987)
142. T.S. Huang, Digital holography. Proc. IEEE. **59**(9), 1335–1346 (1971)
143. A.W. Lohmann, D. Paris, Binary Fraunhofer holograms, generated by computer. Appl. Opt. **6**(10), 1739–1748 (1967)
144. W.-H. Lee, Binary computer-generated holograms. Appl. Opt. **18**(21), 3661–3669 (1979)
145. U. Schnars, W.P.O. Jüptner, Digital recording and numerical reconstruction of holograms, Meas. Sci. Technol. **13**(9), R85–R101 (2002)
146. N. Bokor, Z. Papp, Monte Carlo method in computer holography, Opt. Eng. **36**(4), 1014–1020 (1997)
147. C. Ren, J. Zhou, W. Gao, Four-channel self-focus computer-generated hologram. Appl. Opt. **36**(34), 8844–8847 (1997)
148. O. Falkenstörfer, N. Lindlein, T. Keinonen, J. Schwider, Diffraction-limited holographic lenses in dichromated gelatine, Opt. Eng. **35**(7), 2026–2033 (1996)
149. F. Wyrowski, Diffractive optical elements: iterative calculation of quantized, blazed phase structures. JOSA A **7**(6), 961–969 (1990)
150. L.O. Bubulac, W.E. Tennant, J.G. Pasko, L.J. Kozlowski, M. Zandian, M.E. Motamedi, R.E. DeWames, J. Bajaj, N. Nayar, W.V. McLevige, N.S. Gluck, R. Melendes, D.E. Cooper, D.D. Edwall, J.M. Arias, R. Hall, A.I. D'Souza, et al., High performance SWIR HgCdTe detector arrays. J. Electron. Mat. **26**(6), 649–655 (1997)
151. M.E. Motamedi, W.E. Tennant, H.O. Sankur, R. Melendes, N.S. Gluck, S.Park, J.M. Arias, J. Bajaj, J.G. Pasko, W.V. McLevige, M. Zandian, R.L. Hall, P.D. Richardson, Micro-optic integration with focal plane arrays. Opt. Eng. **36**(5), 1374–1381 (1997)
152. H.A. Macleod, *Thin-Film Optical Filters* (Institute of Physics Publishing, Bristol, 2001)
153. J. Dobrowolski, A. Tikhonravov, M.K. Trubetskov, B.T. Sullivan, P. Verly, Optimal single-band normal-incidence antireflection coatings. Appl. Opt. **35**(4), 644–658 (1996)
154. J.A. Dobrowolski, Optical properties of films and coatings, Handb. Opt. **1,** 42.1–42.130 (1995)

155. L. Young, Synthesis of multiple antireflection films over a prescribed frequency band. JOSA **51**(9), 967–974 (1961)
156. D.J. Aiken, High performance anti-reflection coatings for broadband multi-junction solar cells. Sol. Energy Mater. Sol. Cells **64**(4), 393–404 (2000)
157. U.B. Schallenberg, U. Schulz, N. Kaiser, Multicycle AR coatings: a theoretical approach pp. 357–366
158. J. Dobrowolski, P. Panchhi, M. High, Antireflection coatings designed for two different infrared substrates. Appl. Opt. **35**(1), 102–105 (1996)
159. L. Epstein, The design of optical filters. J. Opt. Soc. Am. A **42**(11), 806–808 (1952)
160. J. Dobrowolski, D. Poitras, P. Ma, H. Vakil, M. Acree, Toward perfect antireflection coatings: numerical investigation. Appl. Opt. **41**(16), 3075–3083 (2002)
161. R. Jacobsson, Light reflection from films of continuously varying refractive index. Prog. Opt. **5**, 247–286 (1966)
162. W.H. Southwell, Gradient-index antireflection coatings. Opt. Lett. **8**(11), 584–586 (1983)
163. O. Stenzel, *The Physics of Thin Film Optical Spectra: An Introduction* (Springer, Berlin, 2005)
164. W.H. Southwell, Pyramid-array surface-relief structures producing antireflection index matching on optical surfaces. JOSA A **8**(3), 549–553 (1991)
165. J.-Q. Xi, M.F. Schubert, J.K. Kim, E.F. Schubert, M. Chen, S.-Y. Lin, W. Liu, J.A. Smart, Optical thin-film materials with low refractive index for broadband elimination of Fresnel reflection. Nat. Photonics **1**(3), 176–179 (2007)
166. M.J. Minot, Single-layer, gradient refractive index antireflection films effective from 0.35 to 2.5 μ. JOSA **66**(6), 515–519 (1976)
167. C. Striemer, P. Fauchet, Dynamic etching of silicon for broadband antireflection applications. Appl. Phys. Lett. **81**(16), 2980–2982 (2002)
168. X. Li, J. Gao, L. Xue, Y. Han, Porous Polymer Films with Gradient-Refractive-Index Structure for Broadband and Omnidirectional Antireflection Coatings. Adv. Funct. Mater. **20** (2), 259–265 (2010)
169. D.H. Raguin, G.M. Morris, Antireflection structured surfaces for the infrared spectral region. Appl. Opt. **32**(7), 1154–1167 (1993)
170. L.D. Landau, E.M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1984)
171. A. Lakhtakia, B. Michel, W.S. Weiglhofer, The role of anisotropy in the Maxwell Garnett and Bruggeman formalisms for uniaxial particulate composite media. J. Phys. D **30**(2), 230–240 (1997)
172. D.A.G. Bruggeman, Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen, I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper ausisotropen Substanzen. Ann. Phys. Lpz **24**, 636–679 (1935)
173. T.G. Mackay, A. Lakhtakia, A limitation of the Bruggeman formalism for homogenization. Opt. Commun. **234**(1–6), 35–42 (2004)
174. T.K. Gaylord, W. Baird, M. Moharam, Zero-reflectivity high spatial-frequency rectangular-groove dielectric surface-relief gratings. Appl. Opt. **25**(24), 4562–4567 (1986)
175. S.M. Rytov, Electromagnetic properties of stratified media (in Russian). J. Exp. Theor. Phys. **29**, 605–608 (1955)
176. S.M. Rytov, Electromagnetic properties of a finely stratified medium. Sov. Phys. JETP **2**, 466–475 (1956)
177. L. Mandel, E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, 1995)
178. A. Yariv, P. Yeh, Electromagnetic propagation in periodic stratified media. II. Birefringence, phase matching, and x-ray lasers. J. Opt. Soc. Am. **67**(4), 438–447 (1977)
179. P. Yeh, A. Yariv, C.-S. Hong, Electromagnetic propagation in periodic stratified media. I. General theory. J. Opt. Soc. Am. **67**(4), 423–438 (1977)
180. P. Yeh, *Optical Waves in Layered Media* (Wiley VCH, Weinheim, 1988)

181. P. Lalanne, D. Lemercier-Lalanne, On the effective medium theory of subwavelength periodic structures. J. Mod. Opt. **43**(10), 2063–2085 (1996)
182. J.M. Dos Santos, L.M. Bernardo, Antireflection structures with use of multilevel subwavelength zero-order gratings. Appl. Opt. **36**(34), 8935–8938 (1997)
183. M. Moharam, T.K. Gaylord, Diffraction analysis of dielectric surface-relief gratings. JOSA **72**(10), 1385–1392 (1982)
184. T.K. Gaylord, M. Moharam, Analysis and applications of optical diffraction by gratings. Proc. IEEE **73**(5), 894–937 (1985)
185. C.-H. Chang, R. Heilmann, R. Fleming, J. Carter, E. Murphy, M. Schattenburg, T. Bailey, J. Ekerdt, R. Frankel, R. Voisin, Fabrication of sawtooth diffraction gratings using nanoimprint lithography. J. Vac. Sci. Technol. B: Microelectron. Nanometer Struct. **21**(6), 2755–2759 (2003)
186. H. Chen, S. Chuang, C. Lin, Y. Lin, Using colloidal lithography to fabricate and optimize sub-wavelength pyramidal and honeycomb structures in solar cells. Opt. Express **15**(22), 14793–14803 (2007)
187. H.Y. Koo, D.K. Yi, S.J. Yoo, D.Y. Kim, A Snowman-like Array of Colloidal Dimers for Antireflecting Surfaces. Adv. Mat. **16**(3), 274–277 (2004)
188. C. Battaglia, C.-M. Hsu, K. Söderström, J. Escarré, F.-J. Haug, M. Charrière, M. Boccard, M. Despeisse, D.T. Alexander, M. Cantoni, Light trapping in solar cells: can periodic beat random? ACS Nano **6**(3), 2790–2797 (2012)
189. P. Spinelli, E. Ferry, J. Van De Groep, M. Van Lare, A. Verschuuren, I. Schropp, A. Atwater, et al., Plasmonic light trapping in thin-film Si solar cells. J. Opt. **14**(2) 024002 (2012)
190. P. Vukusic, J.R. Sambles, Photonic structures in biology. Nature **424**(6950), 852–855 (2003)
191. A.R. Parker, H.E. Townley, Biomimetics of photonic nanostructures. Nat. Nanotech. **2**(6), 347–353 (2007)
192. P.B. Clapham, M.C. Hutley, Reduction of lens reflexion by the ‘moth eye’ principle. Nature **244**(5414), 281–282 (1973)
193. A.B. Harker, J.F. DeNatale, Diamond gradient index moth-eye antireflection surfaces for LWIR windows, pp. 261–267
194. Y.-F. Huang, S. Chattopadhyay, Y.-J. Jen, C.-Y. Peng, T.-A. Liu, Y.-K. Hsu, C.-L. Pan, H.-C. Lo, C.-H. Hsu, Y.-H. Chang, Improved broadband and quasi-omnidirectional anti-reflection properties with biomimetic silicon nanostructures. Nat. Nanotech. **2**(12), 770–774 (2007)
195. C. Wang, S. Yu, W. Chen, C. Sun, Highly efficient light-trapping structure design inspired by natural evolution. Sci. Rep. **3**, 1025 (2013)
196. J. Gittleman, E. Sichel, H. Lehmann, R. Widmer, Textured silicon: A selective absorber for solar thermal conversion. Appl. Phys. Lett. **35**(10), 742–744 (1979)
197. W.W. Wenas, A. Yamada, M. Konagai, K. Takahashi, Textured ZnO thin films for solar cells grown by metalorganic chemical vapor deposition. Jpn. J. Appl. Phys. Part 2: Lett. **30**(3), 441–443 (1991)
198. J. Krč, M. Zeman, O. Kluth, F. Smole, M. Topič, Effect of surface roughness of ZnO: Al films on light scattering in hydrogenated amorphous silicon solar cells. Thin Solid Films **426** (1), 296–304 (2003)
199. C. Rockstuhl, S. Fahr, K. Bittkau, T. Beckers, R. Carius, F.J. Haug, T. Söderström, C. Ballif, F. Lederer, Comparison and optimization of randomly textured surfaces in thin-film solar cells. Opt. Express **18**(19), A335–A342 (2010)
200. J. Haynos, J. Allison, R. Arndt, A. Meulenberg, The comsat non-reflective silicon solar cell: a second generation improved cell. in *International Conference on Photovoltaic Power Generation, Hamburg* (1974), p. 146
201. F. Restrepo, C. Backus, On black solar cells or the tetrahedral texturing of a silicon surface. IEEE Trans. Electron Devices **23**(10), 1195–1197 (1976)
202. M. Adachi, M. Anantram, K. Karim, Core-shell silicon nanowire solar cells. Sci. Rep. **3**, 27 (2013)

203. V. Srivastav, R. Pal, H.P. Vyas, Overview of etching technologies used for HgCdTe. Opto-el. Rev. **13**(3), 197–211 (2005)
204. K. Vynck, M. Burresi, F. Riboli, D.S. Wiersma, Photon management in two-dimensional disordered media. Nat. Mater.**11**(12), 1017–1022 (2012)
205. W. Cai, V. Shalaev, *Optical Metamaterials: Fundamentals and Applications* (Springer, Dordrecht, 2009)
206. S.A. Ramakrishna, T.M. Grzegorczyk, *Physics and Applications of Negative Refractive Index Materials* (SPIE Press Bellingham, WA & CRC Press, Taylor & Francis Group, Boca Raton, 2009)
207. P.W. Milonni, *Fast Light, Slow Light and Left-Handed Light* (Taylor & Francis, Abingdon, Oxford, 2004)
208. V.G. Veselago, The electrodynamics of substances with simultaneously negative values of mu and epsilon. Sov. Phys. Usp. **10**(4), 509–514 (1968)
209. J.B. Pendry, A.J. Holden, D.J. Robbins, W.J. Stewart, Magnetism from conductors and enhanced nonlinear phenomena. IEEE T. Microw. Theory **47**(11), 2075–2084 (1999)
210. R.A. Shelby, D.R. Smith, S. Schultz, Experimental verification of a negative index of refraction. Science **292**(5514), 77–79 (2001)
211. J.B. Pendry, Negative refraction makes a perfect lens. Phys. Rev. Lett. **85**(18), 3966–3969 (2000)
212. H. Chen, C.T. Chan, P. Sheng, Transformation optics and metamaterials. Nat. Mater. **9**(5), 387–396 (2010)
213. J.B. Pendry, D. Schurig, D.R. Smith, Controlling electromagnetic fields. Science **312**(5781), 1780–1782 (2006)
214. W. Cai, U.K. Chettiar, A.V. Kildishev, V.M. Shalaev, Optical cloaking with metamaterials. Nat. Photonics **1**(4), 224–227 (2007)
215. D. Schurig, J.J. Mock, B.J. Justice, S.A. Cummer, J.B. Pendry, A.F. Starr, D.R. Smith, Metamaterial electromagnetic cloak at microwave frequencies. Science **314**(5801), 977–980 (2006)
216. A.I. Fernández-Domínguez, S.A. Maier, J.B. Pendry, Collection and concentration of light by touching spheres: a transformation optics approach. Phys. Rev. Lett. **105**(26) 266807 (2010)
217. Z. Jacob, L.V. Alekseyev, E. Narimanov, Optical hyperlens: far-field imaging beyond the diffraction limit. Opt. Express **14**(18), 8247–8256 (2006)
218. D. Bria, B. Djafari-Rouhani, A. Akjouj, L. Dobrzynski, J. Vigneron, E. El Boudouti, A. Nougaoui, Band structure and omnidirectional photonic band gap in lamellar structures with left-handed materials. Phys. Rev. E **69**(6), 066613 (2004)
219. L.-G. Wang, H. Chen, S.-Y. Zhu, Omnidirectional gap and defect mode of one-dimensional photonic crystals with single-negative materials. Phys. Rev. B **70**(24), 245102 (2004)
220. J. Li, L. Zhou, C. Chan, P. Sheng, Photonic band gap from a stack of positive and negative index materials. Phys. Rev. Lett. **90**(8), 083901 (2003)
221. I.V. Shadrivov, A.A. Sukhorukov, Y.S. Kivshar, Beam shaping by a periodic structure with negative refraction. Appl. Phys. Lett. **82**(22), 3820–3822 (2003)
222. M. Maksimović, Z. Jakšić, Modification of thermal radiation by periodical structures containing negative refractive index metamaterials. Phys. Lett. A **342**(5–6), 497–503 (2005)
223. N. Engheta, An idea for thin subwavelength cavity resonators using metamaterials with negative permittivity and permeability. IEEE Anten. Wirel. Propag. Lett. **1**, 10–13 (2002)
224. M. Maksimović, Z. Jakšić, Emittance and absorptance tailoring by negative refractive index metamaterial-based Cantor multilayers. J. Opt. A-Pure Appl. Opt. **8**(3), 355–362 (2006)
225. M. Maksimović, Z. Jakšić, Transmission spectra of Thue-Morse multilayers containing negative index metamaterials. Acta Phys. Pol. A **112**(5), 1049–1054 (2007)
226. L. Feng, X.-P. Liu, M.-H. Lu, Y.-F. Chen, Phase compensating effect in left-handed materials. Phys. Lett. A **332**(5), 449–455 (2004)

227. H. Cory, C. Zach, Wave propagation in metamaterial multi-layered structures. Microw. Opt. Techn. Let. **40**(6), 460–465 (2004)
228. J. Valentine, S. Zhang, T. Zentgraf, E. Ulin-Avila, D.A. Genov, G. Bartal, X. Zhang, Three-dimensional optical metamaterial with a negative refractive index. Nature **455**(7211), 376–379 (2008)
229. N.I. Landy, S. Sajuyigbe, J.J. Mock, D.R. Smith, W.J. Padilla, Perfect metamaterial absorber, Phys. Rev. Lett. **100**(20), 207402 (2008)
230. N. Liu, M. Mesch, T. Weiss, M. Hentschel, H. Giessen, Infrared perfect absorber and its application as plasmonic sensor. Nano Lett. **10**(7), 2342–2348 (2010)
231. C. Hu, L. Liu, Z. Zhao, X.n. Chen, X. Luo, Mixed plasmons coupling for expanding the bandwidth of near-perfect absorption at visible frequencies. Opt. Express. **17**(19), 16745–16749 (2009)
232. Y. Wang, T. Sun, T. Paudel, Y. Zhang, Z. Ren, K. Kempa, Metamaterial-plasmonic absorber structure for high efficiency amorphous silicon solar cells. Nano Lett. **12**(1), 440–445 (2011)
233. K. Aydin, V.E. Ferry, R.M. Briggs, H.A. Atwater, Broadband polarization-independent resonant light absorption using ultrathin plasmonic super absorbers. Nat. Commun. **2**, 517 (2011)
234. K.B. Alici, E. Ozbay, Photonic Metamaterial Absorber Designs for Infrared Solar Cell Applications, p. 77721B
235. K.B. Alici, A.B. Turhan, C.M. Soukoulis, E. Ozbay, Optically thin composite resonant absorber at the near-infrared band: a polarization independent and spectrally broadband configuration. Opt. Express **19** (15), 14260–14267 (2011)
236. J. Sun, L. Liu, G. Dong, J. Zhou, An extremely broad band metamaterial absorber based on destructive interference. Opt. Express **19**(22), 21155–21162 (2011)
237. Z. Jakšić, Z. Djurić, Cavity enhancement of Auger-suppressed detectors: A way to background-limited room-temperature operation in 3–14 μm range. IEEE J. Sel. Top. Quant. Electr. **10**(4), 771–776 (2004)
238. A.D. Rakić, A.B. Djurišić, J.M. Elazar, M.L. Majewski, Optical properties of metallic films for vertical-cavity optoelectronic devices. Appl. Opt. **37**(22), 5271–5283 (1998)
239. R.A. Paquin, Properties of metals. in *Handbook optics 2, M. Bass*, ed by E.V. Stryland, D. Williams et al. (New York, Mcgraw Hill, 1995), pp. 35.31–35.78
240. E.D. Palik (ed.), *Handbook of Optical Constants of Solids* (Academic Press, Orlando, 1998)
241. J.B. Pendry, Calculating photonic band structure. J. Phys.: Condens. Matter **8**(9), 1085 (1996)
242. A. Yariv, P. Yeh, *Optical waves in crystals* (Wiley, New York, 1984)
243. C.C. Katsidis, D.I. Siapkas, General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference. Appl. Opt. **41**(19), 3978–3987 (2002)
244. J. Joannopoulos, R. Meade, J. Winn, *Photonic Crystals: Molding the Flow of Light*, 2nd edn. (Princeton University Press, Princeton, 2008)
245. N. Matuschek, F.X. Kartner, U. Keller, Theory of double-chirped mirrors. IEEE J. Sel. Top. Quantum Electron. **4**(2), 197–208 (1998)
246. R. Szipöcs, K. Ferencz, C. Spielmann, F. Krausz, Chirped multilayer coatings for broadband dispersion control in femtosecond lasers. Opt. Lett. **19**(3), 201–203 (1994)
247. Y. Fink, J.N. Winn, S. Fan, C. Chen, J. Michel, J.D. Joannopoulos, E.L. Thomas, A dielectric omnidirectional reflector. Science **282**(5394), 1679–1682 (1998)
248. D.N. Chigrin, A.V. Lavrinenko, D.A. Yarotsky, S.V. Gaponenko, All-dielectric one-dimensional periodic structures for total omnidirectional reflection and partial spontaneous emission control. Lightwave Technol. J. **17**(11), 2018–2024 (1999)
249. J.N. Winn, Y. Fink, S. Fan, J. Joannopoulos, Omnidirectional reflection from a one-dimensional photonic crystal. Opt. Lett. **23**(20), 1573–1575 (1998)

250. Z. Djurić, Z. Jakšić, D. Randjelović, T. Danković, W. Ehrfeld, A. Schmidt, Enhancement of radiative lifetime in semiconductors using photonic crystals. Infrared Phys. Technol. **40**(1), 25–32 (1999)
251. M.S. Ünlü, S. Strite, Resonant cavity enhanced photonic devices. J. Appl. Phys. **78**(2), 607–639 (1995)
252. G. Hass, H. Schroeder, A. Turner, Mirror coatings for low visible and high infrared reflectance. JOSA **46**(1), 31–35 (1956)
253. J.C. Fan, F.J. Bachner, G.H. Foley, P.M. Zavracky, Transparent heat-mirror films of TiO_2/Ag/TiO_2 for solar energy collection and radiation insulation. Appl. Phys. Lett. **25**(12), 693–695 (1974)
254. T. Eisenhammer, M. Lazarov, M. Leutbecher, U. Schöffel, R. Sizmann, Optimization of interference filters with genetic algorithms applied to silver-based heat mirrors. Appl. Opt. **32** (31), 6310–6315 (1993)
255. S.M. Vuković, Z. Jakšić, I.V. Shadrivov, Y.S. Kivshar, Plasmonic crystal waveguides. Appl. Phys. A **103**(3), 615–617 (2011)
256. Z. Liu, J. Xu, Z. Lin, Omnidirectional reflection from a slab of uniaxially anisotropic negative refractive index materials. Opt. Commun. **240**(1), 19–27 (2004)
257. H. Da, C. Xu, Z. Li, Omnidirectional reflection from one-dimensional quasi-periodic photonic crystal containing left-handed material. Phys. Lett. A **345**(4), 459–468 (2005)
258. H.H. Barrett, S.F. Jacobs, Retroreflective arrays as approximate phase conjugators. Opt. Lett. **4**(6), 190–192 (1979)
259. J. Yao, X. Yang, X. Yin, G. Bartal, X. Zhang, Three-dimensional nanometer-scale optical cavities of indefinite medium. Proc. Natl Acad. Sci. **108**(28), 11327–11331 (2011)
260. Y.G. Ma, C. Ong, T. Tyc, U. Leonhardt, An omnidirectional retroreflector based on the transmutation of dielectric singularities. Nat. Mater. **8**(8), 639–642 (2009)
261. C. Bohren, D. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983)
262. W.E. Vargas, Optimization of the diffuse reflectance of pigmented coatings taking into account multiple scattering. J. Appl. Phys. **88**(7), 4079–4084 (2000)
263. F.E. Gálvez, P.R. Barnes, J. Halme, H. Míguez, Dye sensitized solar cells as optically random photovoltaic media. Energy Environ. Sci. **7**(2), 689–697 (2014)
264. D.S. Wiersma, P. Bartolini, A. Lagendijk, R. Righini, Localization of light in a disordered medium. Nature **390**(6661), 671–673 (1997)
265. F.J.P. Schuurmans, Light in complex dielectrics, Ph.D. dissertation, University of Amsterdam, Amsterdam, 1999
266. M. Quinten, *Optical Properties of Nanoparticle Systems: Mie and Beyond* (Wiley-VCH, Weinheim, 2011)
267. A. Meulenberg, The sawtooth coverslide-A new means of coupling light into solar cells. J. Energy **1**(3), 151–154 (1977)
268. T.I. Chappell, The V-groove multijunction solar cell. IEEE Trans. Electron Devices **26**(7), 1091–1097 (1979)
269. A. Vujanić, M. Smiljanić, Photodiode quantum efficiency optimization using spherical pits on active surfaces. Microelectron. J. **26**(5), 413–419 (1995)
270. M.A. Green, The path to 25% silicon solar cell efficiency: history of silicon cell evolution. Prog. Photovoltaics: Res. Appl. **17**(3), 183–189 (2009)
271. H.-S. Lee, S.M. Sze, Silicon pin photodetector using internal reflection method. IEEE Trans. Electron Devices **17**(4), 342–347 (1970)
272. D. Redfield, Multiple-pass thin-film silicon solar cell. Appl. Phys. Lett. **25**, 647 (1974)
273. D. Kleppner, Inhibited spontaneous emission. Phys. Rev. Lett. **47**(4), 233–236 (1981)
274. S. Murtaza, K. Anselm, C. Hu, H. Nie, B. Streetman, J. Campbell, Resonant-cavity enhanced (RCE) separate absorption and multiplication (SAM) avalanche photodetector (APD). IEEE Photonics Technol. Lett. **7**(12), 1486–1488 (1995)

275. M.S. Ünlü, B.M. Onat, Y. Leblebici, Transient simulation of heterojunction photodiodes-part II: analysis of resonant cavity enhanced photodetectors. Lightwave Technol. J. **13**(3), 406–415 (1995)

276. J. Pautrat, E. Hadji, J. Bleuse, N. Magnea, Resonant-cavity infrared optoelectronic devices. J. Electron. Mat. **26**(6), 667–672 (1997)

277. J. Goedbloed, J. Joosten, Responsivity of avalanche photodiodes in the presence of multiple reflections. Electron. Lett. **12**(14), 363–364 (1976)

278. E. Ozbay, M. S. Islam, B. Onat, M. Gokkavas, O. Aytur, G. Tuttle, E. Towe, R. Henderson et al., Fabrication of high-speed resonant cavity enhanced Schottky photodiodes. IEEE Photonics Technol. Lett. **9**(5), 672–674 (1997)

279. B. Onat, M. Gokkavas, E. Ozbay, E. Ata, E. Towe, M. Unlu, 100-GHz resonant cavity enhanced Schottky photodiodes. IEEE Photonics Technol. Lett. **10**(5), 707–709 (1998)

280. R. Sidhu, H. Chen, N. Duan, G. Karve, J. Campbell, A. Holmes, GaAsSb resonant-cavity enhanced avalanche photodiode operating at 1.06 μm. Electron. Lett. **40**(20), 1296–1297 (2004)

281. X. Chen, B. Nabet, F. Quaranta, A. Cola, M. Currie, Resonant-cavity-enhanced heterostructure metal-semiconductor-metal photodetector. Appl. Phys. Lett. **80**(17), 3222–3224 (2002)

282. M. Ünlü, K. Kishino, J.-I. Chyi, L. Arsenault, J. Reed, S.N. Mohammad, H. Morkoç, Resonant cavity enhanced AlGaAs/GaAs heterojunction phototransistors with an intermediate InGaAs layer in the collector. Appl. Phys. Lett. **57**(8), 750–752 (1990)

283. M. Ünlü, K. Kishino, J.-I. Chyi, J. Reed, L. Arsenault, H. Morkoc, Wavelength demultiplexing heterojunction phototransistor. Electron. Lett. **26**(22), 1857–1858 (1990)

284. R. Bockstaele, T. Coosemans, C. Sys, L. Vanwassenhove, A. Van Hove, B. Dhoedt, I. Moerman, P. Van Daele, R.G. Baets, R. Annen, Realization and characterization of 8 × 8 resonant cavity LED arrays mounted onto CMOS drivers for POF-based interchip interconnections. IEEE J. Sel. Top. Quantum Electron. **5**(2), 224–235 (1999)

285. E. Hadji, J. Bleuse, N. Magnea, J. Pautrat, 3.2 μm infrared resonant cavity light emitting diode. Appl. Phys. Lett. **67**(18), 2591–2593 (1995)

286. J. Wehner, C. Musca, R. Sewell, J. Dell, L. Faraone, Mercury cadmium telluride resonant-cavity-enhanced photoconductive infrared detectors. Appl. Phys. Lett., **87**(21), 211104–211104-3 (2005)

287. A.M. Green, D.G. Gevaux, C. Roberts, C.C. Phillips, Resonant-cavity-enhanced photodetectors and LEDs in the mid-infrared. Phys. E **20**(3), 531–535 (2004)

288. E. Yablonovitch, Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett. **58**(20), 2059–2062 (1987)

289. J.-M. Lourtioz, H. Benisty, V. Berger, J.-M. Gerard, D. Maystre, A. Tchelnokov, *Photonic Crystals: Towards Nanoscale Photonic Devices*, 2nd edn. (Springer-Verlag, Berlin, 2008)

290. E. Yablonovitch, T. Gmitter, R. Meade, A. Rappe, K. Brommer, J. Joannopoulos, Donor and acceptor modes in photonic band structure. Phys. Rev. Lett. **67**(24), 3380 (1991)

291. D. Smith, R. Dalichaouch, N. Kroll, S. Schultz, S. McCall, P. Platzman, Photonic band structure and defects in one and two dimensions. JOSA B **10**(2), 314–321 (1993)

292. P. Bermel, C. Luo, L. Zeng, L. C. Kimerling, J. D. Joannopoulos, Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals, Opt. Express **15**(25), 16986–17000 (2007)

293. R.C. Schroden, M. Al-Daous, C.F. Blanford, A. Stein, Optical properties of inverse opal photonic crystals. Chem. Mat. **14**(8), 3305–3315 (2002)

294. V. Altschul, E. Finkman, Simple approximation for Fermi energy in nonparabolic semiconductors. Appl. Phys. Lett. **58**(9), 942–944 (1991)

295. Z. Jakšić, Z. Djurić, Z. Djinović, V. Jović, F. Kermendi, O. Jakšić, Analysis of Optical Constants of Experimentally Fabricated Epitaxial HgCdTe with Arbitrary Degree of Carrier Degeneration. Solid State Phenom. B: Diffus. Defect Data **7**, 335–338 (1998)

296. C. Bauer, G. Kobiela, H. Giessen, 2D quasiperiodic plasmonic crystals. Sci. Rep. **2**, 0681.1–0681.6 (2012)
297. Z. Jakšić, Optical metamaterials as the platform for a novel generation of ultrasensitive chemical or biological sensors. in *Metamaterials: Classes, Properties and Applications*, ed by E. J. Tremblay (Hauppauge, Nova Science Publishers, New York, 2010), pp. 1–42
298. P. Drude, *The Theory of Optics* (Dover Publications, Mineola, 2005)
299. S.V. Boriskina, H. Ghasemi, G. Chen, Plasmonic materials for energy: from physics to applications. Mater. Today, **16**(10), 375–386 (2013)
300. S. Franzen, Surface plasmon polaritons and screened plasma absorption in indium tin oxide compared to silver and gold. J. Phys. Chem. C **112**(15), 6027–6032 (2008)
301. P. Avouris, M. Freitag, Graphene photonics, plasmonics, and optoelectronics. IEEE J. Sel. Top. Quant. Electron. **20**(1), 600112 (2014)
302. H.A. Atwater, A. Polman, Plasmonics for improved photovoltaic devices. Nat. Mater. **9**(3), 205–213 (2010)
303. D. Derkacs, S. Lim, P. Matheu, W. Mar, E. Yu, Improved performance of amorphous silicon solar cells via scattering from surface plasmon polaritons in nearby metallic nanoparticles. Appl. Phys. Lett. **89**(9), 093103 (2006)
304. J.A. Schuller, E.S. Barnard, W. Cai, Y.C. Jun, J.S. White, M.L. Brongersma, Plasmonics for extreme light concentration and manipulation. Nat. Mater. **9**(3), 193–204 (2010)
305. A. Aubry, D.Y. Lei, A.I. Fernández-Domínguez, Y. Sonnefraud, S.A. Maier, J.B. Pendry, Plasmonic light-harvesting devices over the whole visible spectrum. Nano Lett. **10**(7), 2574–2579 (2010)
306. T. Trupke, M. Green, P. Würfel, Improving solar cell efficiencies by up-conversion of sub-band-gap light. J. Appl. Phys. **92**(7), 4117–4122 (2002)
307. G. Shvets, Y.A. Urzhumov, Electric and magnetic properties of sub-wavelength plasmonic crystals. J. Opt. A-Pure Appl. Opt. **7**(2), S23–S31 (2005)
308. P. Biagioni, J.-S. Huang, B. Hecht, Nanoantennas for visible and infrared radiation. Rep. Prog. Phys. **75**(2), 024402 (2012)
309. T.W. Ebbesen, H.J. Lezec, H.F. Ghaemi, T. Thio, P.A. Wolff, Extraordinary optical transmission through sub-wavelength hole arrays. Nature **391**(6668), 667–669 (1998)
310. J.B. Pendry, L. Martín-Moreno, F.J. Garcia-Vidal, Mimicking surface plasmons with structured surfaces. Science **305**(5685), 847–848 (2004)
311. J. Rosenberg, R.V. Shenoi, T.E. Vandervelde, S. Krishna, O. Painter, A multispectral and polarization-selective surface-plasmon resonant midinfrared detector. Appl. Phys. Lett. **95** (16), 161101 (2009)
312. A.F. Koenderink, Plasmon nanoparticle array waveguides for single photon and single plasmon sources. Nano Lett. **9**(12), 4228–4233 (2009)
313. P.J. Schuck, D.P. Fromm, A. Sundaramurthy, G.S. Kino, W.E. Moerner, Improving the mismatch between light and nanoscale objects with gold bowtie nanoantennas. Phys. Rev. Lett. **94**(1), 017402 (2005)
314. T. Grosjean, M. Mivelle, F.I. Baida, G.W. Burr, U.C. Fischer, Diabolo nanoantenna for enhancing and confining the magnetic optical field. Nano Lett. **11**(3), 1009–1013 (2011)
315. J. Li, A. Salandrino, N. Engheta, Shaping light beams in the nanometer scale: a yagi-uda nanoantenna in the optical domain. Phys. Rev. B **76**(24), 245403 (2007)
316. E.N. Grossman, J.E. Sauvageau, D.G. McDonald, Lithographic spiral antennas at short wavelengths. Appl. Phys. Lett. **59**(25), 3225–3227 (1991)
317. G. Volpe, G. Volpe, R. Quidant, Fractal plasmonics: subdiffraction focusing and broadband spectral response by a Sierpinski nanocarpet. Opt. Express **19**(4), 3612–3618 (2011)
318. Y. Alaverdyan, B. Seplveda, L. Eurenius, E. Olsson, M. Käll, Optical antennas based on coupled nanoholes in thin metal films. Nat. Phys. **3**(12), 884–889 (2007)
319. B.L. Twu, S.E. Schwarz, Properties of infrared cat-whisker antennas near 10.6 μ. Appl. Phys. Lett. **26**(12), 672–675 (1975)

320. M.W. Knight, H. Sobhani, P. Nordlander, N.J. Halas, Photodetection with active optical antennas, Science **332**(6030), 702–704, (2011)
321. Z. Jakšić, M. Milinović, D. Randjelović, Nanotechnological enhancement of infrared detectors by plasmon resonance in transparent conductive oxide nanoparticles. Strojniski Vestnik/J. Mech. Eng. **58**(6), 367–375 (2012)
322. K.R. Catchpole, A. Polman, Plasmonic solar cells. Opt. Express **16**(26), 21793–21800 (2008)
323. S. Lal, S. Link, N.J. Halas, Nano-optics from sensing to waveguiding. Nat. Photonics **1**(11), 641–648 (2007)
324. T. Ung, L.M. Liz-Marzan, P. Mulvaney, Optical properties of thin films of Au@ SiO_2 particles. J. Phys. Chem. B **105**(17), 3441–3452 (2001)
325. C. Bauer, H. Giessen, Light harvesting enhancement in solar cells with quasicrystalline plasmonic structures. Opt. Express **21**(103), A363–A371 (2013)
326. V.M. Hietala, G. Vawter, T. Brennan, B. Hammons, Traveling-wave photodetectors for high-power, large-bandwidth applications. IEEE Trans. Microwave Theory Tech. **43**(9), 2291–2298 (1995)
327. V. Vickers, Model of Schottky barrier hot-electron-mode photodetection. Appl. Opt. **10**(9), 2190–2192 (1971)
328. M. Nawrocki, Y.G. Rubo, J. Lascaray, D. Coquillat, Suppression of the Auger recombination due to spin polarization of excess carriers and Mn^{2+} ions in the semimagnetic semiconductor Cd_ {0.95} Mn_ {0.05} S. Phys. Review B, **52**(4), R2241 (1995)
329. A. Imamoglu, Y. Yamamoto, Noise suppression in semiconductor pin junctions: transition from macroscopic squeezing to mesoscopic Coulomb blockade of electron emission processes. Phys. Rev. Lett. **70**(21), 3327 (1993)
330. M. De Jong, C. Beenakker, Semiclassical theory of shot-noise suppression. Phys. Rev. B **51** (23), 16867 (1995)
331. T. Morimoto, M. Chiba, G. Kido, A. Tanaka, Stimulated emission due to the magnetoelectric photoeffect in narrow-gap semiconductors at the quantum limit. Semicond. Sci. Tech. **8**(1S), S417 (1993)
332. T. Ashley, C. Elliott, A. White, Infra-red detection using minority carrier exclusion. in *1985 International Technical Symposium/Europe* 1986, pp. 62–68
333. T. Ashley, C.T. Elliott, A.T. Harker, Non-equilibrium modes of operation for infrared detectors. Infrared Phys. **26**(5), 303–315 (1986)
334. Z. Djuric, J. Piotrowski, Room temperature IR photodetector with electromagnetic carrier depletion. Electron. Lett. **26**(20), 1689–1691 (1990)
335. A. Medvid, V. Sotnikov, Thermogradient luminescence. Infrared Phys. **29**(2), 705–707 (1989)
336. T. Ashley, C.T. Elliott, Operation and properties of narrow-gap semiconductor devices near room temperature using non-equilibrium techniques, *Semicond. Sci. Tech.* **6**, 12 C (1991)
337. V.K. Malyutenko, S.S. Bolgov, E.I. Yablonovsky, A new type of IR luminescence. Infrared Phys. **25**(1–2), 115–119 (1985)
338. P. Berdahl, V. Malyutenko, T. Morimoto, Negative luminescence of semiconductors. Infrared Phys. **29**(2–4), 667–672 (1989)
339. C.T. Elliott, N.T. Gordon, A.M. White, Towards background-limited, room-temperature, infrared photon detectors in the 3–13 μm wavelength range. Appl. Phys. Lett. **74**(19), 2881–2883 (1999)
340. J. Singh, *Physics of Semiconductors and Their Heterostructures* (McGraw Hill, New York, 1993)
341. K. Seeger, *Semiconductor Physics* (Springer-Verlag, Wien, 1973)
342. G. Baccarani, M. Rudan, R. Guerrieri, P. Ciampolini, Physical Models for Numerical Device Simulation. Process Device Model. ed. by W. L. Engl (Elsevier, Amsterdam, 1986), pp. 107–158
343. S. Selberherr, *Analysis and Simulation of Semiconductor Devices* (Springer Verlag, Berlin, 1984)

344. A. Anselm, *Introduction to Semiconductor Theory* (Mir Publishers, Moscow, 1981)
345. P. S. Kireev, Semiconductor Physics (Mir, Moscow, 1978)
346. R.S. Popovic, *Hall Effect Devices* (CRC Press, Boca Raton, 2010)
347. B. Nag, *Electron Transport in Compound Semiconductors* (Springer Verlag, Berlin, 1980)
348. R.F. Pierret, G.W. Neudeck, *Advanced semiconductor fundamentals* (Addison-Wesley Reading, Massachusetts, 1987)
349. Y. He, Discretization method for current continuity equation containing magnetic field. IEEE Trans. Electron Devices **39**(4), 820–824 (1992)
350. H. Fujisada, Effects of diffusion current on galvanomagnetic properties in thin intrinsic InSb at room temperature. J. Appl. Phys. **45**(8), 3530–3540 (1974)
351. A. Rogalski, *Infrared Photon Detectors* (SPIE Press, Bellingham, 1995)
352. D. Lile, Generalized photoelectromagnetic effect in semiconductors. Phys. Rev. B **8**(10), 4708 (1973)
353. C. Summers, B. Darling, B. Martin, Computer modeling of carrier transport in (Hg, Cd) Te photodiodes. J. Appl. Phys. **59**(7), 2457–2466 (1986)
354. J. Arthur, W. Bardsley, M. Brown, A. Gibson.Carrier extraction in germanium. Proc. Phys. Soc. Sect. B, **68**(1), 43 1(955)
355. R.W. Lade, A.G. Jordan, On the Static Characteristics of High-Low Junction Devices. Int. J. Electron. **13**(1), 23–34 (1962)
356. T. Ashley, C.T. Elliott, Nonequilibrium devices for infra-red detection. Electron. Lett. **21** (10), 451–452 (1985)
357. T. Ashley, C.T. Elliott, A.M. White, Infra-Red Detection Using Minority Carrier Exclusion, pp. 62–68
358. Z. Djuric, V. Jovic, M. Matic, Z. Jaksic, IR photodetector with exclusion effect and self-filtering n+ layer. Electron. Lett. **26**(13), 929–931 (1990)
359. Z. Djuric, B. Livada, V. Jovic, M. Smiljanic, M. Matic, Z. Lazic, Quantum efficiency and responsivity of InSb photodiodes utilizing the Moss-Burstein effect. Infrared Phys. **29**(1), 1–7 (1989)
360. L.S. Stil'bans, Fizika poluprovodnikov (in Russian) *(*Sovetskoe radio, Moscow, 1967)
361. C.T. Elliott, N.T. Gordon, R. Hall, T. Phillips, C. Jones, B. Matthews, C. Maxey, et al., Metal organic vapor phase epitaxy (MOVPE) grown heterojunction diodes in $Hg_{1-x}Cd_xTe$. in SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation, (1994), pp. 648–657
362. C. Elliott, N. Gordon, R. Hall, T. Phillips, A. White, C. Jones, C. Maxey, N. Metcalfe, Recent results on metalorganic vapor phase epitaxially grown HgCdTe heterostructure devices. J. Electron. Mat. **25**(8), 1139–1145 (1996)
363. T. Ashley, A. Dean, C. Elliott, C. McConville, G. Pryce, C. Whitehouse, Ambient temperature diodes and field-effect transistors in $InSb/In_{1-x}Al_xSb$. Appl. Phys. Lett. **59**(14), 1761–1763 (1991)
364. T. Ashley, C. Elliott, N. Gordon, R. Hall, A. Johnson, G. Pryce, Uncooled $InSb/In_{1-x}Al_xSb$ mid-infrared emitter. Appl. Phys. Lett. **64**(18), 2433–2435 (1994)
365. T. Ashley, C. Elliott, N. Gordon, R. Hall, A. Johnson, G. Pryce, Room temperature narrow gap semiconductor diodes as sources and detectors in the 5–10 μm wavelength region. J. Cryst. Growth **159**(1), 1100–1103 (1996)
366. P. Emelie, J. Phillips, S. Velicu, P. Wijewarnasuriya, Parameter extraction of HgCdTe infrared photodiodes exhibiting Auger suppression. J. Phys. D **42**(23), 234003 (2009)
367. P.Y. Emelie, S. Velicu, C.H. Grein, J.D. Phillips, P.S. Wijewarnasuriya, N.K. Dhar, Modeling of LWIR HgCdTe auger-suppressed infrared photodiodes under nonequilibrium operation. J. Electron. Mat. **37**(9), 1362–1368 (2008)
368. P. Martyniuk, A. Rogalski, HOT infrared photodetectors. Opto-el. Rev. **21**(2), 239–257 (2013)

369. A.M. Itsuno, J.D. Phillips, S. Velicu, Predicted performance improvement of Auger-Suppressed HgCdTe Photodiodes and p-n Heterojunction Detectors. IEEE Trans. Electron Devices **58**(2), 501–507 (2011)
370. M. Kopytko, K. Jóźwikowski, Numerical Analysis of Current-Voltage Characteristics of LWIR nBn and p-on-n HgCdTe Photodetectors. J. Electron. Mat. **42**(11), 3211–3216 (2013)
371. S. Maimon,G. Wicks, nBn detector, an infrared detector with reduced dark current and higher operating temperature. Appl. Phys. Lett. **89(**15), 151109–3 (2006)
372. D.Z. Ting, C.J. Hill, A. Soibel, J. Nguyen, S.A. Keo, J.M. Mumolo, M.C. Lee, B. Yang, et al., Antimonide superlattice barrier infrared detectors, in *Proceedings SPIE 7419, Infrared Systems Photoelectronic Technol IV,* (2009), pp. 74190B–12
373. A. Itsuno, J. Phillips, S. Velicu, Design and modeling of HgCdTe nBn detectors. J. Electron. Mat. **40**(8), 1624–1629 (2011)
374. A.M. Itsuno, P.Y. Emelie, J.D. Phillips, S. Velicu, C.H. Grein, P.S. Wijewarnasuriya, Arsenic diffusion study in HgCdTe for low p-type doping in auger-suppressed photodiodes. J. Electron. Mat. **39**(7), 945–950 (2010)
375. P. Martyniuk, W. Gawron, A. Rogalski, Theoretical modeling of HOT HgCdTe Barrier detectors for the Mid-Wave Infrared Range. J. Electron. Mat. **42**(11), 3309–3319 (2013)
376. C.T. Elliott, A.M. White, Multiple heterostructure photodetector, US Patent 5,068,524, 1991
377. T. Ashley, C. Elliott, A. White, G. Crimes, A. Harker, Near-ambient-temperature bipolar transistor in cadmium mercury telluride. Electron. Lett. **23**(24), 1280–1281 (1987)
378. W. Van Roosbroeck, The transport of added current carriers in a homogeneous semiconductor. Phys. Rev. **91**(2), 282 (1953)
379. M. Kurata, *Numerical Analysis for Semiconductor Devices* (Lexington Books, Massachusetts, 1982)
380. A.M. White, Auger suppression and negative resistance in low gap PIN diode structures. Infrared Phys. **26**(5), 317–324 (1986)
381. A.M. White, Negative resistance with Auger suppression in near-intrinsic, low-bandgap photo-diode structures. Infrared Phys. **27**(6), 361–369 (1987)
382. J. Piotrowski, A. Jóźwikowska, K. Jóźwikowski, R. Ciupa, Numerical analysis of longwavelength extracted photodiodes. Infrared Phys. **34**(6), 565–572 (1993)
383. T. Skauli, H. Steen, T. Colin, P. Helgesen, S. Løvold, C.T. Elliott, N.T. Gordon, T.J. Phillips, A.M. White, Auger suppression in CdHgTe heterostructure diodes grown by molecular beam epitaxy using silver as acceptor dopant. Appl. Phys. Lett. **68**(9), 1235–1237 (1996)
384. T. Ashley, C. Elliott, Operation and properties of narrow-gap semiconductor devices near room temperature using non-equilibrium techniques. Semicond. Sci. Tech. **6**(12C), C99 (1991)
385. C. Elliott, N. Gordon, R. Hall, T. Phillips, C. Jones, A. Best, 1/f noise studies in uncooled narrow gap $Hg_{1-x}Cd_xTe$ non-equilibrium diodes. J. Electron. Mat. **26**(6), 643–648 (1997)
386. C. Elliott, N. Gordon, T. Phillips, H. Steen, A. White, D. Wilson, C. Jones, C. Maxey, N. Metcalfe, Minimally cooled heterojunction laser heterodyne detectors in metalorganic vapor phase epitaxially grown $Hg_{1-x}Cd_xTe$. J. Electron. Mat. **25**(8), 1146–1150 (1996)
387. G. Nash, M. Ashby, J. Lindle, N. Gordon, W. Bewley, J. Meyer, J. Giess, L. Haworth, T. Ashley, Long wavelength infrared negative luminescent devices with strong Auger suppression. J. Appl. Phys. **94**(11), 7300–7304 (2003)
388. K. Jóźwikowski, A. Rogalski, A. Jóźwikowska, Numerical modeling of fluctuation phenomena in semiconductors and detailed noise study of mid-wave infrared HgCdTe-heterostructure devices. J. Electron. Mat. **31**(7), 677–682 (2002)
389. K. Jozwikowski, R. Sewell, C. Musca, J. Dell, L. Faraone, Noise modeling in HgCdTe heterostructure devices. J. Appl. Phys. **94**(10), 6541–6548 (2003)
390. H. Schaake, M. Kinch, D. Chandra, F. Aqariden, P. Liao, D. Weirauch, C.-F. Wan, R. Scritchfield, W. Sullivan, J. Teherani, High-operating-temperature MWIR detector diodes. J. Electron. Mat. **37**(9), 1401–1405 (2008)

391. M. Ashby, N. Gordon, C. Elliott, C. Jones, C. Maxey, L. Hipwood, R. Catchpole, Investigations into the source of 1/f noise in $Hg_xCd_{1-x}Te$ diodes. J. Electron. Mat. **33**(6), 757–760 (2004)
392. C. Jones, N. Metcalfe, A. Best, R. Catchpole, C. Maxey, N. Gordon, R. Hall, T. Colin, T. Skauli, Effect of device processing on 1/f noise in uncooled, Auger-suppressed CdHgTe diodes. J. Electron. Mat. **27**(6), 733–739 (1998)
393. H. Suhl, W. Shockley, Concentrating holes and electrons by magnetic fields. Phys. Rev. **75** (10), 1617 (1949)
394. S. Sze, Semiconductor Sensors. (Wiley, New York, 1994)
395. A. Akopyan, Z. Gribnikov, K. Guga, M. Malozovskii, V. Malyutenko, Otricatel'naya differencial'naya provodimost' pri magnitokoncentracionnom effekte v nesobstvennom poluprovodnike (in Russian). Fizika i Tekhnika Poluprovodnikov **13**(11), 2111–2119 (1979)
396. K.Y. Guga, Y.M. Malozovskiy, V.K. Malyutenko, N-ODP v usloviyah magnitokoncentra cionnogo effekta pri nelineinoi rekombinaciii nositelei toka. UDK **10**, 1858–1861 (1982)
397. V. Malyutenko, G. Teslenko, Photoconductors under transverse sweep-out conditions. Infrared Phys. **25**(1), 337–342 (1985)
398. V. Malyutenko, G. Teslenko, I. Boiko, Transverse sweep-out effects in insb in crossed electric and magnetic fields. Infrared Phys. **13**(4), 337–339 (1973)
399. I.L. Bazhenov, B.L. Gel'mont, V.I. Ivanov-Omskiy, A.I. Izhin, V.A. Smirnov, Fotolyuminescenciya tverd'ih rastvorov $Cd_{0.4}Hg_{0.6}Te$. Fizika i Tekhnika Poluprovodnikov **24**(1), 93–97 (1990) (in Russian)
400. Z. Jakšić, Z. Djurić, Optimized high-frequency performance of Auger-suppressed magnetoconcentration photoconductors. Microelectron. J. **31**(11–12), 981–990 (2000)
401. P. Berdahl, Galvanomagnetic luminescence and the quantum efficiency of radiative recombination of InSb. J. Appl. Phys. **63**(12), 5846–5858 (1988)
402. A. Akopyan, S. Bolgov, A. Savcenko, O. Salyuk, Magnitokoncentracionnij effekt v neodnorodno legirovannih poluprovodnikah (in Russian). Fizika i Tekhnika Poluprovodnikov **24**(10), 1875–1878 (1990)
403. V.K. Malyutenko, A.P. Medvid', V.A. Mis'nik, Control Over the Spectral Sensitivity of InSb Photoelectric Detectors. Sov. Phys. and Tech. of Semicond. **10**(4), 646–651 (1976)
404. Z. Djuric, J. Piotrowski, Dember IR photodetectors. Solid State Electron. **34**(3), 265–269 (1991)
405. Z. Djurić, A. Vujanić, Z. Jakšić, Spectral characteristics of high temperature IR photodetectors with electromagnetic carrier depletion. Infrared Phys. Technol. **35**(4), 585–591 (1994)
406. J.M. Kahn, Wireless infrared communications. IEEE Proc. **85**(2), 265–298 (1997)
407. D.J. Heatley, D.R. Wisely, I. Neild, P. Cochrane, Optical wireless: the story so far. IEEE Commun. Mag. **36**(12), 72–74, 79–82 (1998)
408. E. Korevaar, I. Kim, B. McArthur, Debunking the recurring myth of a magic wavelength for free-space optics. 155–166
409. H. Manor, S. Arnon, Performance of an optical wireless communication system as a function of wavelength. Appl. Opt. **42**(21), 4285–4294 (2003)
410. M. Achour, Free-space optics wavelength selection: 10 μm versus shorter wavelengths. J. Opt. Netw. **2**(6), 127–143 (2003)
411. P. Corrigan, R. Martini, E.A. Whittaker, C. Bethea, Quantum cascade lasers and the Kruse model in free space optical communication. Opt. Express **17**(6), 4355–4359 (2009)
412. D.P. Hutchinson, R.K. Richards, All-weather long-wavelength infrared free space optical communications. J. Opt. Fiber Comm. Res. **4**(3), 214–224 (2007)
413. D.A. Rockwell, G.S. Mecherle, "Wavelength selection for optical wireless communications systems", in ITCom. Int. Symp. Convergence IT Commun. **2001**, 27–35 (2001)
414. Z. Jakšić, Nonequilibrium photodetectors for wireless infrared communications. in *Proceedings 10th Telecommunications Forum TELFOR 2002*, 26–29, Belgrade, 2002, pp. 654–657
415. Z. Jakšić, Z. Djurić, Extraction Photodiodes with Auger Suppression for All-Weather Free-Space Optical Communication. Electronics **8**(1), 30–32 (2004)

Index

A
Absorption coefficient, 13, 19, 39, 43, 45, 105
Absorption cross-section, 126
Acceptor, 29, 114, 164
Accumulation, 152
Active region, 13
Ambipolar diffusion, 13, 182
Ambipolar transport, 181
Anderson localization, 103
Anti-barrier, 152
Antireflection, 41
Antireflection films, 71, 74
Antireflection structures (ARS), 44, 49, 70, 77, 80, 82
AR coatings, 72, 78, 83
Auger, 22, 39, 42, 130, 136, 151, 154, 165, 175, 179, 181, 195, 201, 206, 207, 211
Auger 1, 24, 27, 137, 157
Auger 7, 27
Auger generation, 129, 136, 165, 186, 188
Auger lifetime, 24–27, 134, 219
Auger lifetime, intrinsic, 25
Auger recombination, 17, 22, 23, 130, 167
Auger suppression, 132, 139, 140, 155, 174, 178, 183, 206
Avalanche photodiodes, 105

B
Babinet principle, 124
Background limited infrared photodetector. *See* BLIP, 2
Background radiation, 33
Backscattering, 102
Bandgap narrowing, 28
Bandgap, 39
Barrier, 152, 180
Beamsplitting, 69
Beattie, 22, 23, 28, 134, 136, 183
Bichromated gelatin, 54
Bio-inspired design, 83
Biomimetics, 83
Bionics, 83
BIRD detector, 180
Black silicon, 82
BLIP, 40, 105, 114, 118, 129, 134, 135, 165, 175
Bloch, 96, 116
Bloch vector, 116
Bloch wave functions, 25
Bolometer, 6
Boltzmann equation, 140, 142
Bose-Einstein distribution, 21
Bowtie nanoantenna, 124
Bragg condition, 70
Bragg equation, 78
Bragg gratings, 82
Bragg mirrors, 94, 95, 114
Brewster's angle, 103
Bromine-methanol, 53
Bruggeman, 77
Bulk micromachining, 104
Burstein-Moss, 27

C
Capture constants, 31
Carrier mobility, 15
Cavity enhancement, 91, 114
CCCH, 22, 23, 34
Chebyshev polynomial, 96
Chemical micromachining, 53, 54
Chemical vapor deposition, 75, 85
Chemomechanical polishing, 54
CHHL, 22, 23, 34
Chirped mirrors, 98
Clausius-Mossotti, 77
Computer-generated holography (CGH), 54, 61, 68, 69
Concentration ratio, 47

© Springer International Publishing Switzerland 2014

Z. Jakšić, *Micro and Nanophotonics for Semiconductor Infrared Detectors*,
DOI 10.1007/978-3-319-09674-2

Conduction band, 22, 114
Confocal systems, 51
Conic concentrator, 59
Continuity equation, 37, 141, 154, 159, 162, 181
Cooling requirements, 5, 11
Core-shell nanoparticle, 126
Coulomb interaction, 22, 25, 26, 28
Coupling efficiency, optical, 48
CPC, 55, 57
CPC trough, 57
Critical length, 204
Cross-antenna, 124
Cryogenic cooling, 11, 41, 129, 130, 228
Current sensitivity, 216
Current-voltage characteristics, 168, 188, 209
Cutoff wavelength, 13, 18, 39, 134, 139

D
D*f* product, 12, 15
Damping factor, 121
Dark current, 12, 34, 37, 133, 171
DBR, 94, 115, 118
Debye length, 25
Debye temperature, 136
Degenerate semiconductors, 23, 27
Dember effect, 226
Dember magnetoconcentration photodetector, 226
Dember voltage, 227
Density of states, 21, 22, 157, 192
Density of states, optical, 46, 121
Depletion region, 14, 15, 203
Designer plasmons, 122, 123
Detector active area, 12
Detector bandwidth, 2
Detector element, 1
Detector response, 5, 16, 125, 217
Detector thickness, 39
Dielectric mirrors, 92, 97, 116
Dielectric permittivity, 18, 19, 78, 141
Diffraction, 93
Diffraction efficiency, 69
Diffraction equation, 78
Diffraction patterns, 125
Diffractive gratings, 77
Diffractive lens, 66
Diffractive optical element (DOE), 47, 52, 53, 61–63, 68, 70, 103
Diffractive order, 78
Diffuse reflection, 92, 101, 102
Diffusion constant, 14
Diffusion length, 176
Dispersion relation, plasmonic, 121
Distributed Bragg mirrors, 94
Donor, 29, 114, 164
Drude model, 121, 125
Dry etching, 53, 82, 85
Dynamic resistance, 39, 219

E
Effective conductivity, 146
Effective medium theory, 77, 87
Effective refractive index, 80, 86
Einstein relation, 150
Electrochemical etching, 76
Electron beam, 82
Electron-beam, 53
Elliott, 158, 191
Emissivity, 111
Etching, 103
Evanescent field, 123
Evaporation, 75
Exclusion contact, 157
Exclusion length, 162, 166, 206
Exclusion, 131, 152, 156, 158, 161–163, 166, 168, 170, 173, 177–180, 184
Extraction, 131, 152, 175, 177, 178, 180, 184, 185, 188, 190
Extraordinary optical transmission (EOT), 122, 123
Extrinsic detector, 12

F
Fabry-Perot, 105
Fast light, 86, 121
Fermi function, 30
Fermi level, 30, 143, 146
Figures of merit, 2
Filling factor, 80
Flicker noise, 36, 133, 175, 228
Floquet, 96, 116
Focused ion beam, 82, 90
Focusing optics, 47
Form birefringence, 79
Free-space optics, 228
Frequency response, 217
Fresnel lens, 68, 69
Fresnel phase plate, 67
Fresnel reflectance, 96
Fresnel reflector, 60
Fresnel zone plate (FZP), 64–66
Fresnel-Kirchhoff integral, 65

G
Galvanomagnetic, 196
Gauss law, 140
Gaussian distribution, 64
Generalized photonic IR detector, 40
Generation term, 201
Generation-recombination, 16, 17, 21, 39
Generation-recombination rate, 35, 37, 142
Geometric concentration ratio, 47
g-r noise, 37, 129
Gradient refractive index, 51, 80
Gradient-index materials, 76
Gradient-index thin films, 76
Graphene, 121
Grating vector, 78
Grayscale DOE, 61
GRIN, 51, 53

H
Hall mobility, 145, 148
Hartree-Fock, 25, 142
Haze, 85
Heavy holes band, 22
Heterodyne detector, 11, 195
HgCdTe, 85, 53, 70, 71, 78, 107, 119, 134, 146, 148, 157, 164, 184
High-reflection surfaces, 91
Holographic lithography, 54, 82
Holographic optical element (HOE), 48, 51, 53, 61, 68, 69
Holography, 68
Homojunction, 193
Hot electrons, 130, 175, 206
HOT, 196
Hotspot, 90, 123
Hyperbolic concentrator, 59
Hyperlens, 87

I
Imaging detector, 1
Immersion, 41, 48, 49
Immersion lens, 49
Impact ionization, 22, 26, 129
Impedance matching, 87
Incident photon flux density, 12
Infrared detector, 1, 122
Infrared radiation, 1, 158
Injection molding, 52
Injection, carrier, 152
InSb, 22, 53, 178, 219
Interference, 93, 105
Interference films, 71
Interference mirrors, 93, 94
Intrinsic detector, 12
Intrinsic noise, 34
Intrinsic semiconductor, 22
Inverse lasers, 130
Inverted opal, 114
Invisibility cloak, 87
Ion etching, 104
Ion implantation, 82
Ion milling, 53, 82
Irradiance gain, 47
Isotype junction, 157, 168, 182

J
Johnson noise, 35, 37, 38, 40
Johnson–Nyquist noise, 35
Josephson detector, 8
Joule heating, 135, 139, 223

K
Kane matrix element, 138
Kane’s model, 19
Kinoform, 68, 69
Kronig–Penney model, 23, 116

L
Lambertian, 46, 49, 102, 103
Lambertian limit, 45
Landsberg, 22, 23, 134, 136, 183
Leakage currents, 194
LED, 105
Left-handed light, 86, 121
Left-handed materials, 87
Lift-off, 82
Light concentrator, 47, 55
Light holes band, 22
Light localization, 44
Light management, 41, 43
Light pipe, 91, 111, 112
Light trapping, 42, 45, 90, 91
Liouville-Von Neumann expression, 142
Localization, 103
Lord Rayleigh, 77
Lorentz force, 139, 144, 149, 221
Lorentzian spectrum, 195
Lorenz-Lorentz equation, 77
LWIR, 1, 11, 41, 51, 75, 135, 228

M
Macroporous media, 102

Magnetoconcentration, 131, 196–198, 201, 204, 207, 209, 210, 215, 217, 218, 220, 222
Magnetoconcentration-extraction device, 224, 225
Malyutenko, 198
Maxwell equations, 77, 115, 140
Maxwell-Boltzmann, 200
Maxwell-Garnett, 77
MBE, 195
MEMS, 51, 197
Mercury cadmium telluride, 16, 19, 29, 82, 134, 166, 168, 170, 173, 178, 183, 218
Mesoscopic, 104
Metal mirrors, 94
Metal nanoparticles, 126
Metal-dielectric mirrors, 100
Metalorganic vapor phase epitaxy, 194
Metamaterial-containing multilayers, 88
Metamaterials, 86, 87, 121
Microcavity, 110
Microfabrication, 49
Microgrinding, 54
Microlens, 49
Microreflectors, 55
Microsystem technologies, 7, 49
Mie scattering, 121, 123
Minority carriers, 129, 176
MOEMS, 94
Molding, 52
MOM diode, 8
Monte-Carlo method, 70, 103
Moth eye, 83, 84
MSM detector, 10
Multilayer, 72
Multilevel Fresnel phase plate, 67
Multiple scattering, 102
MWIR, 1, 41, 51, 75, 135

N
Nanoantennas, 123
Nanocomposites, 120
Nanodimer, 123
Nanoholes, 125
Nanoparticle chain, 123
Nanoparticles, 120, 121, 125
Nanorice, 126
Nanorod antenna, 123
Nanoshells, 126
Nanosphere lithography, 82
Nanosphere, 123
Nanovoids, 121
Narrow-bandgap semiconductors, 16, 33, 131, 140
nBn, 180
Needle optimization, 74
Negative index metamaterials, 87
Negative luminescence, 133, 198
Negative refractive index, 87
Night vision, 1
NIR, 1
1/f noise, 36, 195
Noise, 34, 174, 191
Noise current, 15, 37, 38, 170, 212, 214
Noise suppression, 42, 131
Non-directing junction, 152
Nonequilibrium, 42, 129–132, 135, 139, 152, 181
Nonequilibrium depletion, 134, 221
Nonequilibrium detector, 41, 132, 133, 140, 149
Nonequilibrium methods, 129, 130
Nonimaging concentrators, 47, 54
Nonimaging optics, 47
NWIR, 11

O
Omnidirectional mirrors, 98
Omnidirectional, 76, 82, 98
Omnidirectional reflectivity, 93, 98
Optical antenna, 123
Optical concentrator, 41, 43, 47, 51
Optical generation, 32, 155
Optical grating, 61, 78
Optical interference coatings, 95
Optical lithography, 53
Optical parametric converter, 8
Optical path, 39
Optical trapping, 44
Overlap integral, 23

P
Parabolic reflector, 57
PBG, 112, 114, 115, 117, 131
PCE, 114, 115
PEM, 197
Pendry, 87, 122, 123
Penetration depth, 94
Penrose tiling, 128
Perfect black surfaces, 89
Perfect lens, 87
Photochemical etching, 54
Photoconductivity, 12, 36

Photoconductor, 10, 16, 212
Photodetector, 43, 47, 105, 107, 110, 114, 115, 118, 121, 125, 129, 130, 196, 205, 210, 222
Photodiode, 12, 35, 158, 175, 179
Photoelectric gain, 12, 37, 39, 190, 212
Photoelectromagnetic, 10, 197
Photoemissive vacuum detectors, 8
Photolithographic mask, 82
Photolithography, 52, 90, 104, 114
Photomasks, 82
Photomultiplicator, 8
Photon absorption, 21, 40
Photon management, 41, 43, 45
Photon recycling, 20
Photonic bandgap, 88, 97, 110, 112, 114
Photonic crystal enhancement, 92, 112, 114
Photonic crystal, 116, 93, 97, 98, 112, 114, 115, 118, 119
Photonic detector, 9, 11, 15
Photoresist, 53
Photovoltaic, 36, 38
Photovoltaic detector, 10, 14, 15, 111
p-i-n photodiodes, 105
Planar technology, 52
Planck law, 18
Plasma frequency, 121–123
Plasmonic enhancement, 128
Plasmonic material, 120
Plasmonic nanoparticles, 90, 125, 126
Plasmonics, 120, 121
p-n junction, 14
Poisson equation, 140, 155
Porous silica, 76
Poynting vector, 87
Principle of detailed balance, 20, 46
Projection printing, 53
Pseudoimmersion, 48
PV detector. *See* Photovoltaic detector, 10
Pyroelectric detector, 16

Q
QDIP, 9
Quadratic programming, 74
Quantum efficiency, 12–15, 39, 41, 43, 104, 106, 107
Quarterwave stack, 88, 94, 98
Quarter-wavelength, 74
Quasicrystal, 128
Quasi-Fermi level, 150
Quasiperiodic, 120
QWIP, 9

R
Radiance theorem, 46
Radiative generation, 18
Radiative lifetime, 19, 20, 22, 118, 119, 134, 135, 202
Radiative recombination, 17, 20, 207
Radiative shield, 110, 111
Random antireflection profile, 84
Ray optics, 46, 47
Ray optics limit, 45
RCE, 105, 107, 131
Reabsorption, 20, 135
Reactive-ion etching, 53, 82
Recombination, 16, 152, 155, 167, 182, 188, 207
Recombination rate, 33, 209, 212
Redshifting, 125
Reflection coefficient, 13, 92, 113
Reflective concentrator, 47, 54, 56
Reflective losses, 49
Refractive index, 18, 40, 72, 74, 76, 78, 95, 102, 106, 114, 119, 120
Refractive lenses, 48
Refractive optics, 47
Reliability, 5
Resonant cavity, 105, 131
Resonant cavity enhancement, 92, 105, 107
Response time, 5, 12, 15, 172, 217, 219
Responsivity, 5, 109, 212
Retroreflection, 101
RIE, 53
Room temperature photodetectors, 213
Roosbroeck-Shockley's model, 193
Rytov, 78

S
Scattering, 85, 101
Scattering cross-section, 125, 126
Scattering sponges, 102
Schottky diodes, 105
Schottky noise, 35
Sensitivity, 170, 193, 214
Shockley, 17, 149
Shockley-Read, 31, 167, 180, 188, 200, 201, 207
Shockley-Read lifetime, 31
Shockley-Read recombination, 17, 29
Shockley-Read-Hall. *See* Shockley-Read, 29
Shot noise, 35
Skin depth, 94
Slow light, 86, 121
Solar cells, 47, 125

Solar energy concentration, 58
Specific detectivity, 2, 11, 12, 13, 15, 41, 109, 171, 193, 212, 214, 216
Spectral noise, 221
Spectral selectivity, 6
Specular reflection, 92, 93
Spin split-off band, 22
Split ring resonators, 90
Spontaneous radiation, 111
Spoof plasmons, 123
SPP, 120, 125
Sputtering, 53, 82, 85
Step-down refractive index, 73
Stop band, 88, 98
Subsurface scatterers, 85
Subwavelength, 104, 121
Subwavelength plasmonic crystals, 122
Suhl effect, 196
Superabsorber, 89
Superconcentrator, 87
Superisolators, 111
Superlens, 87
Superluminal group velocities, 86
Surface plasmon polaritons, 44, 120–122
Surface recombination rate, 13, 14, 201, 202
Sweepout, 15, 161, 174

T
Thermal detectors, 6
Thermal equilibrium, 18
Thermal melting, 51
Thermal noise, 35, 42
Thermoelectric coolers, 218
Total internal reflection, 103, 104
Transfer matrix technique, 95, 97, 117
Transformation optics, 101, 121
Transit time, 173
Transmission coefficient, 99
Trap, 30
Trumpet concentrator, 59

U
Uncooled detector, 109
Unipolar BIRD, 180
Urbach band tailing, 19
UV replication, 52

V
Vacuum tube, 8
Valence band, 114
Van Roosbroeck, 17, 149
Van Roosbroeck model, 140, 154, 156
Van Roosbroeck-Shockley model, 170
VCSEL, 131
V-groove, 104
Voltage responsivity, 221
Volume hologram, 68
V-trough, 59

W
Wet etching, 82, 85
Winston collector, 55, 57
Winston cone, 54, 55, 60

Y
Yablonovitch, 45, 46
Yagi-Uda nanoantenna, 124

Z
Zero-index band gap, 88
Zincblende, 140, 141